TOWSON UNIVERSITY

OFFICE OF GRADUATE STUDIES

BIOINFORMATICS TOOLS AND ANALYSIS TO IDENTIFY CANDIDATE LETHAL GENES IN ROOT-KNOT NEMATODE (RKN) AND ITS COMPARISON WITH SOYBEAN CYST NEMATODE (SCN) INFECTION IN SOYBEAN

By

Ahmed I. Ismail

A dissertation in partial fulfillment of the requirement for the degree Doctor of Science in Information Technology.

Department of Computer and Information Science

Towson University

Baltimore, MD 21252

August 2013

TOWSON UNIVERSITY

OFFICE OF GRADUATE STUDIES

DISSERTATION APPROVAL PAGE

This is to certify that the dissertation prepared by AHMED ISMAIL entitled "BIOINFORMATICS TOOLS AND ANALYSIS TO IDENTIFY CANDIDATE LETHAL GENES IN ROOT KNOT NEMATODE (RKN) AND ITS COMPARISON WITH SOYBEAN CYST NEMATODE (SCN) INFECTION IN SOYBEAN" has been approved by the thesis committee as satisfactorily completing the dissertation requirements for the degree Doctor of Science in Information Technology.

Nadim Alif — 8/12/2013
Chair, Dissertation Committee — **Date**

Benjamin T Martin — 8/12/2013
Committee Member — **Date**

[signature] — 8/12/13
Committee Member — **Date**

[signature] — 8/13/2013
Committee Member — **Date**

Janet V DiRosa — 8/19/13
Dean of Graduate Studies — **Date**

Acknowledgment

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family and wife.

I would like to express my deepest gratitude to my advisor, Dr. Nadim Alkharouf, for his excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research. I would like to thank Dr. Benjamin F. Matthews, for his assistance, guidance, patiently corrected my writing, and for providing me with the microarray data to be analyzed in our experiment. Special thanks go to Dr. Hammel and Dr. Hong, who were willing to participate in my final defense committee and for their valuable discussions and accessibility.

I would like to thank Omar Darwish, who as a good friend was always willing to help and give his best suggestions. It would have been a lonely lab without him.

I would also like to thank my parents, younger brother, and younger sister. They were always supporting me during the long years of my education and encouraging me with their best wishes.

Finally, I would like to thank my wife. She was always there cheering me up and stood by me through the good times and bad.

Abstract

BIOINFORMATICS TOOLS AND ANALYSIS TO IDENTIFY CANDIDATE LETHAL GENES IN ROOT KNOT NEMATODE (RKN) AND ITS COMPARISON WITH SOYBEAN CYST NEMATODE (SCN) INFECTION IN SOYBEAN.

Ahmed I. Ismail

Root-Knot nematode (RKN; *Meloidogyne spp.*) is one of the most devastating obligate parasites that infect the roots of thousands of plant species. In the United States, the cost of the losses that these parasites cause annually is estimated to be about $8 billion. The RKN genome sequence is not well characterized or fully sequenced. We sought to identify genes that are lethal at any stage of RKN development. These genes could act as targets for RNAi constructs or gene knockout experiments. To that end we compared RKN gene sequences to the well-characterized and fully sequenced genome of the free-living nematode *C. elegans* (*Caenorhabditis elegans*). The primary database for *C. elegans* (www.wormbase.org) contains a huge repository of RNAi and mutation experiments on *C. elegans* genes. We conducted a comparative genomics analysis, comparing RKN genes of six species: *Meloidogyne arenaria*, *M. chitiwoodi*, *M. hapla*, *M. incognita*, *M. javanica*, and *M. paranaensis* to *C. elegans* via WormBase. The RKN sequences were obtained from nematode.net. Our analysis yielded many candidate lethal genes between *Meloidogyne spp.* and *C.elegans*. Some of these genes have already been tested experimentally and indeed found to be lethal to RKN. We also created a relational web based database that links the RKN genes and proteins to their *C. elegans* homologs. The database provides

search capabilities for scientists in the field. The database was created using SQL-Server 2008. The web based user interface was created using ASP.NET.

Using microarray analysis we examined the expression of soybean (Glycine max) genes in roots infected by the RKN, Meloidogyne incognita at 12 days and 10 weeks after infection and similarly soybean roots infected by SCN at 8 days after infection. Gene expression was monitored using the Affymetrix Soybean GeneChip containing 37,500 G. max probe sets. A comparative microarray analysis was done between RKN and SCN using bioinformatics tools such as MATLAB (R2010b) Bioinformatics toolbox and Blast2go. The goal was to identify the common genes that were differentially expressed (induced or suppressed) as a result of the soybean infection by RKN and SCN. Our results showed that there are major differences in the gene expression profile between the soybean infected by SCN and the soybean infected by RKN. Where only 135 genes were shown to be induced and 82 genes were shown to be suppressed as the common altered genes between soybean roots infected by RKN (Meloidogyne incognita) and similarly soybean roots infected by SCN.

Table of Contents

Contents

List of Tables

List of figures

List of Abbreviations

RKN: Root-Knot Nematode

EST: Expressed Sequence Tag

SCN: Soybean Cyst Nematode

Dpi: Days post infection

Wpi: Weeks post infection

PM: Perfect Match

MM: Mismatch

GO: Gene Ontology

RNAi: RNA interference

LCM: Laser Capture Microdissection

BLAST: Basic Local Alignment Search Tool

RKNLDB: Root-Knot Nematode Lethal Database

gi: gene identifiers

Chapter 1

Specific Aims and Objectives

"For every five animal species on the planet, four are nematodes" (Platt 1994). The majority of animals and plants have at least one parasitic nematode species that is specifically tailored to make use of the food and resources that the host species represents.

Plant-parasitic nematodes decrease the product of the world's primary foods by an average of 12.3% (Dawar *et al*. 2008) and Root-Knot Nematodes **(**RKN**)** (*Meloidogyne spp*.) have been classified to be the biggest contributors to these losses. Parasitic nematodes infect more than 2000 plant species. Therefore, controlling these parasites is necessary to maintain good crop production.

RKN are obligatory parasites that infect the roots of thousands of plant species, and cannot live independently from their host. They are dispersed worldwide. While some species have several different races, this specific genus includes more than 60 species. There are four major *Meloidogyne* species that are important worldwide, *M. javanica*, *M. arenaria*, *M. incognita*, and *M. hapla*. (Eisenback and Triantaphyllou 1991).

Vegetable crops that are grown in warm and humid conditions undergo rigorous damages from RKN. Damage from RKN have led to consequences such as: weak growth, a decrease in the quality and quantity of the crop, and reduction of immunity

to other natural epidemics (e.g. drought, other diseases). However, if a large amount of RKN harm occurs, this can result in complete crop loss. The reason for this is that roots damaged by nematodes do not maximize the usage of water and fertilizers, hence, contributing to further losses for the cultivator (Olsen 2000).

The goal of this study is to apply bioinformatics and data mining tools to analyze RKN ESTs (Expressed Sequence Tags) to identify the lethal genes and their pathways. The organism will not be able to survive if these lethal genes either silenced or mutated which will cause organism's death. Since the model free-living nematode "*Caenorhabditis elegans*" has been the subject of intensive study and little is known about the other nematodes, we will compare the DNA and ESTs content of the major six species of *Meloidogyne* we chose with the protein of those of *C.elegans* in order to discover common lethal genes that exist in both nematodes. When the lethal genes in any of our six species of RKN "*Meloidogyne spp.*" are discovered, these genes will be used to find homology for these genes among all RKN species, so one gene can target all RKN species.

Our first objective is to supply details on the implementation of comparing both *C.elegans* and *Meloidogyne spp.* aiming to identify if any of our *Meloidogyne* **s**ix species has candidate lethal genes that are capable of controlling their activities in the plant soil, and consequently, saving crops and money.

Our second objective is to understand the effects of infection of soybean roots by *M. incognita* to represent one of the smallest and recently discovered genomes among nematode and SCN as the most devastating pathogen of soybean. Monitoring the differential expressed genes and demonstrating the numerous changes in gene

expression occur in roots and in syncytial cells in soybean roots infected by SCN and in giant cells of roots infected by RKN. Also, we are trying to find the common altered genes expression between them so that some of these genes may be used as candidates for developing plants resistance to RKN and SCN through over-expression or silencing.

Our third objective is to develop a relational database to store and manage the huge amount of data generated by the previous two experiments. The database will be created using SQLServer2008 and it will be housed at the computer science department at Towson University in Towson, MD.

Our fourth objective is to make that database available to the public via the World Wide Web; the web site will be built using ASP.NET. Web based applications and tools will be developed to query the database from the World Wide Web. Also, I will link my local database with WormBase (http://www.wormbase.org/), Nematode (www.nematode.net), and pubMed (http://www.ncbi.nlm.nih.gov/pubmed/) to see which of these genes produced an embryonic or postembryonic lethal phenotype in C.elegans if mutated or silenced.

The remainder of the dissertation is divided into three major sections. The first section includes all the processes we performed to Identify lethal genes and their pathways in RKN targeting RNAi constructs or gene knockout experiments by using the EST analysis of RKN and by conducting a comparative genomics analysis, comparing RKN to C. elegans. The second section contains the analysis of the two microarray experiments of soybean roots infected by RKN and SCN starting from importing the microarray raw data until identifying which genes are differentially

expressed and which genes are the common altered genes between those two experiments. In the last section we declared that we developed a relational database to store data generated by the experiments from the previous two sections and all their related information. The process of designing the web based user interface to query our database was described in that section to facilitate our data analysis. The dissertation ends with a chapter outlining conclusions and discussing potential future work.

Chapter 2

Identifying Lethal Genes in RKN

2.1 Background and Significance

RKN (*Meloidogyne spp*.) are considered to be key pathogens of vegetables throughout both domestically and globally, which impact both the quantity and quality of profitable yields. Scanning from horticultural, ornamental and vegetable crops, to field crops, Meloidogyne arises in 23 of 43 crops that are scientifically recorded as having plant-parasitic nematodes of chief caliber (Stirling *et al*. 1992). When a situation occurs where RKN infect the deep-root of the crops, controlling it will be hard and solutions will be limited. The cycle begins when matured females lay hundreds of eggs; after these eggs hatch, they relocate to a favorable plant root and begin feeding. These hatched eggs damage plants by implanting a thin tube, a stylet, into the roots and then sucking the sap. This, in turn, causes a decline in the size and efficiency of the root systems.

Figure (1a) is depicting the Scanning electron micrograph of a second-stage RKN juvenile propped up against a human hair. An RKN gall is the most dramatic symptom that is formed as a result of nematode feeding, large galls or "knots" throughout the root system of infected plants. Severe infections result in reduced yields on numerous crops (see figure 1b). Root galls vary in size and shape depending on the type of plant, nematode population levels, and species of root-knot nematode present in the soil.

Figure 1a&b: Plant roots infected by RKN (Pictures are courtesy of Nemapix Jounal of ematological Images, vol. 1 &2, J.D. Eisenback & Ulrich Zunke, eds).

RKN are carried by a large spectrum of hosts; all major field crops, fruit trees, most vegetable crops, ornamental plants, and weeds. Only Two *Meloidogyne spp.* genomes have been sequenced and annotated recently and completely (M. hapla & M. incognita). The common host plants for ***M. hapla*** are strawberry, bean, carrot, cucumber, dill, eggplant, endive, okra, pea, pepper, tomato, potato, and onion. Moreover, the host plants for *M. incognita* are cabbage, cantaloupe, carrot, lettuce, corn, pumpkin, spinach, squash, and watermelon (Widmer *et al.* 1999).

2.1.1 Damage

The annual global agricultural losses of RKN have been estimated to be $157 billion (Abad *et al.* 2008). Carrots are the most sensitive crop; Carrot loss has been estimated by as much as 45% in commercial fields, while the weight loss of onions has been estimated by as much as 70% in commercial fields, while the infection by this nematode can reduce the lettuce weight by 26% in the field (Widmer *et al.* 1999).

2.1.2 Life Cycle

RKN go through the majority of their life cycle within the roots hosting them. They typically survive and live in soil during either the egg stage or the second stage larvae. Mature females of RKN leave a minimum of 1000 eggs in a gelatin-like medium, which can be monitored as being attached to the extending posterior end of the females on the root surface. This pouch prevents eggs dehydration (see Figure 2).

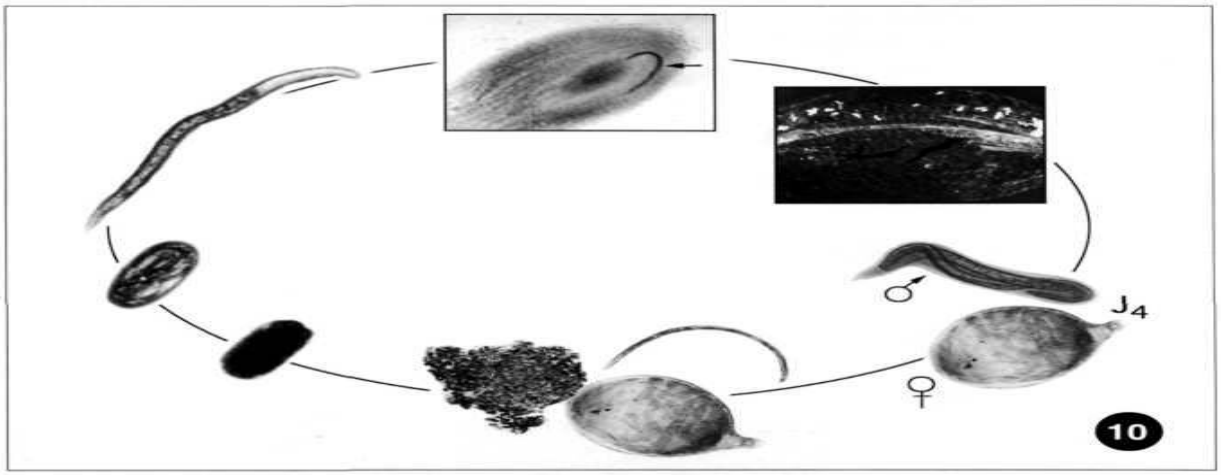


Figure 2: Life cycle of the Root Knot Nematode (RKN).

Source: (http://vegetablemdonline.ppath.cornell.edu/factsheets/RootKnotNematode.htm)

After hatching, in the infective second stage of their lifestyle, the juveniles navigate the soil to search for roots of available host plants. It is customary that the juveniles pierce host roots just behind the root tip region and institute their permanent feeding sites and giant cells in the vascular tissues of the root. The giant cells supply the inactive nematodes nutrients, which continue to feed, enlarge, and molt three times. Around the feeding sites, root cells are induced to grow and create galls (knots) and often cause more root formation as well as branching of the main root. The entire lifecycle may be circulated in 17 to 57 days depending upon the host and soil

temperature. The distribution of nematodes among fields can occur by means of vegetative plant parts, irrigation water, and soil infested with eggs or larvae which comply with farm implements, animals, or humans (Widmer *et al*. 1999).

When RKN infects plant roots, there is a sophisticated interactive relationship with the host cell. Within the nematode's esophageal gland cells, there are different gene products that are expressed to help the nematode establish a feeding site from the host's cells. Some of these proteins secreted by the nematode are injected into the host cell and cause modification of the cell wall (Davis *et al.* 2004).

2.1.3 ESTs and Genome of RKN

In addition to being vital in gene discovery and gene sequence determination, ESTs (Expressed Sequence Tag) may be utilized to measure gene expression (the process of converting gene DNA into mRNA which serves as the template for protein synthesis). The number of ESTs has grown rapidly, around 65.9 million of the ESTs are now available in public databases (e.g. GenBank 18/6/2010, all species). ESTs represent portions of expressed genes because these clones contain DNA that corresponds to mRNA. The ESTs may be present in the database in two different forms: a cDNA/mRNA sequence or in the form of the reverse complement of the mRNA, the template strand. EST and complementary DNA (cDNA) sequences present direct proof for all the transcripts sampled and they are currently the most significant resources for transcription exploration (Hass *et al*. 2003). ESTs used to be short between 200–800 nucleotide bases in length, in an original state, arbitrarily selected single-pass sequence reads derived from cDNA libraries. High-throughput ESTs can

be made at low cost from either the 50 or 30 side of a cDNA clone to get a closer look into the transcriptional active regions in any organism. A widespread growth occurred in creation and collection of ESTs data in the public databases for myriad organisms. Now ESTs enable gene discovery, gene structure identification, and complement genome annotation and sequencing. Also it can be used as another choice in sequencing the genome of many organisms.

We have used EST analysis of RKN ESTs to find candidate lethal genes. Two complete *Meloidogyne* genomes have simultaneously been sequenced and annotated. Other *Meloidogyne spp*. have been sequenced to varying degrees of coverage. The recent completion of the two RKN (*M. hapla* & *M. incognita*) genomes opens the way for a comparative genomic approach to identify these lethal genes so that we can control this parasitic distribution. The sequencing projects revealed that *M. hapla* encodes approximately 14,200 genes 54Mbp genome. This genome represents the smallest nematode genome yet completed and the smallest metazoan genome characterized so far. *M. incognita* encodes approximately 19,200 genes 86Mbp genome (David *et al*. 2009).

According to the latest update of the (NCBI) National Center for Biotechnology Information genomic database, the number of ESTs that has been sequenced is 73,343 for *Meloidogyne spp*, they are distributed as follows: *M. hapla* (24452), *M. incognita* (20334), *M. chitwoodi* (12218), *M. javanica* (7587), *M. arenaria* (5042), and *M. paranaensis* (3710).

In the coming chapters we describe the computational methods and tools used to identify candidate lethal genes. We describe the public database and warehouse that

stores EST data. We also describe the web based applications and tools that were developed to query the database from the web.

2.2 Methodology of development

Simplifying the search for candidate lethal genes would require the creation of a database that compares ESTs or nucleotides sequence information from an uncharacterized genome with that of a well characterized genome. One of the excellent ways to obtain the candidate lethal genes that are lethal at any stage of the RKN development is to compare RKN ESTs and proteins to the well-studied and sequenced genome of *C.elegans*. We chose *C.elegans* because unlike RKN, it is well-characterized and there are numerous knockout and RNAi (**b**iological process in which RNA molecules inhibit gene expression) studies have been done on it, making it a perfect species for comparison. In our experiment I performed pairwise sequence alignments (using BlastN, BlastP, and BlastX) of the RKN unigene sequences versus the *C.elegans* sequences to identify potential lethal genes between the RKN and *C.elegans*.

We generated a flow diagram (Figure 3) that depicts the analysis pipeline that will be developed for the identification of the lethal genes of *Meloidogyne spp*.

Figure 3: RKNLDB Flow diagram.

2.2.1 ESTs and Genomic Analysis

We downloaded all known EST, protein, and genomic sequences for *C.elegans* from (http://www.wormbase.org/). WormBase is an international consortium of biologists and computer scientists dedicated to providing the research community with accurate, current, accessible information concerning the genetics, genomics and biology of *C. elegans* and related nematodes. There were 364064 EST sequences and 22863 genomic sequences as of August 2010. The sequences were prepared and processed according to my database pipeline (Fig.3). We created a local *C.elegans* database from the downloaded sequences, and using SQLServer2008 to create that database. We also downloaded all known EST and genomics sequences for six of the RKN (*Meloidogyne.spp*) species from (http://www.nematode.net), Nematode.net is the home page of the parasitic nematode EST project at The Genome Institute at Washington University in St. Louis. The site was established in 2000 as a component of the NIH-NIAID grant "A Genomic Approach to Parasites from the Phylum Nematoda". While Nematode.net started as a project site, over the years it became a community resource dedicated to the study of parasitic nematodes. The site is the property of Washington University (Martin *et al*. 2009). There were a total of 73,343 EST sequences and 26,067 genomic sequences for the six species as of August 2010 these numbers are divided in (Table 1).

Species	EST Sequences	Genomic Sequences
Meloidogyne hapla	24454	6667
Meloidogyne incognita	20334	5598
Meloidogyne chitwoodi	12218	3578
Meloidogyne javanica	7587	4827
Meloidogyne arenaria	5042	3241
Meloidogyne paranaensis	3710	2156

Table1: all known EST and genomic sequences of RKN (Source:http//www.Nematode.net as of September 2010).

We imported these assembled ESTs into our local database. The consensus sequences for the contigs and the sequences of the singletons were then blasted (**B**asic **L**ocal **A**lignment **S**earch **T**ool, or BLAST, is an algorithm that is used to compare primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences) against the local *C.elegans* database. Customized Perl scripts were written to reformate and parse the results of the blast searches. These scripts extracted the best hits (**t**he highest similar sequences between RKN and *C. elegans*) from the blast results according to some of the parameters such as the E-value, bit score, and identity values. The parsed results were then stored in the local database. SQL scripts were written to query the database for genes with high similarity to *C.elegans*. E-value of 1*E*-5 was chosen as the cut-off value of the *E*-value (**t**he probability of having same similar sequences by chance) parameter.

2.2.2 RNAi Phenotypes discovery

We linked our local database to WormBase server (http://www.wormbase.org/) using WBGeneID, which was one of the primary entities in our local database trying to get more information about our best hits. Customized Perl script was written to extract all the WBGeneIDs and then we submitted them to the WormBase database. As a result for our submission we obtained important information about all the RNAi experiments for each one of the WBGeneIDs in our local database, each RNAi experiment has a unique identification number, the results were then stored back into our local database. RNAi technology allows genome-function screening and is broadly used in the identification of genes associated with specific biological phenotypes (Zhang 2011). Thus, we started looking for identifying all the phenotypes associated with all the RNAi experiments in our local database and also to discover which of these genes produce lethal phenotypes in *C.elegans* if mutated or silenced according to (Alkharouf *et al*. 2006). Another customized Perl script was written to extract all the RNAiIDs and then we linked them back to WormBase, but this time we used one of the tools on the WormBase server that is called WormMart tool.

WormMart (http://caprica.caltech.edu:9002/biomart/martview/) provides easy and interactive access to data. In particular, WormMart enables users to retrieve sequences and genome annotations using a variety of filters to constrain retrieved results. One of the filters included is to retrieve a list of transcription factors which when mutated or targeted by RNAi cause lethal phenotype. Data can be returned in a variety of formats, including HTML, text, FASTA, and even Microsoft Excel.

WormMart is the WormBase implementation of BioMart. In all of our work we were using the latest database version (WS220) on WormBase server by August 2010.

2.3 Results

We sought to identify conserved genes that potentially were lethal at any stage of the *Meloidogyne.spp* development. A reasonable way to obtain such candidate lethal genes was to compare a well-studied and sequenced genome of *C. elegans* with that of the EST database for *Meloidogyne.spp*. We developed a pipeline to acquire and process nucleotide sequences from *Meloidogyne.spp* and *C.elegans* (Fig. 3).

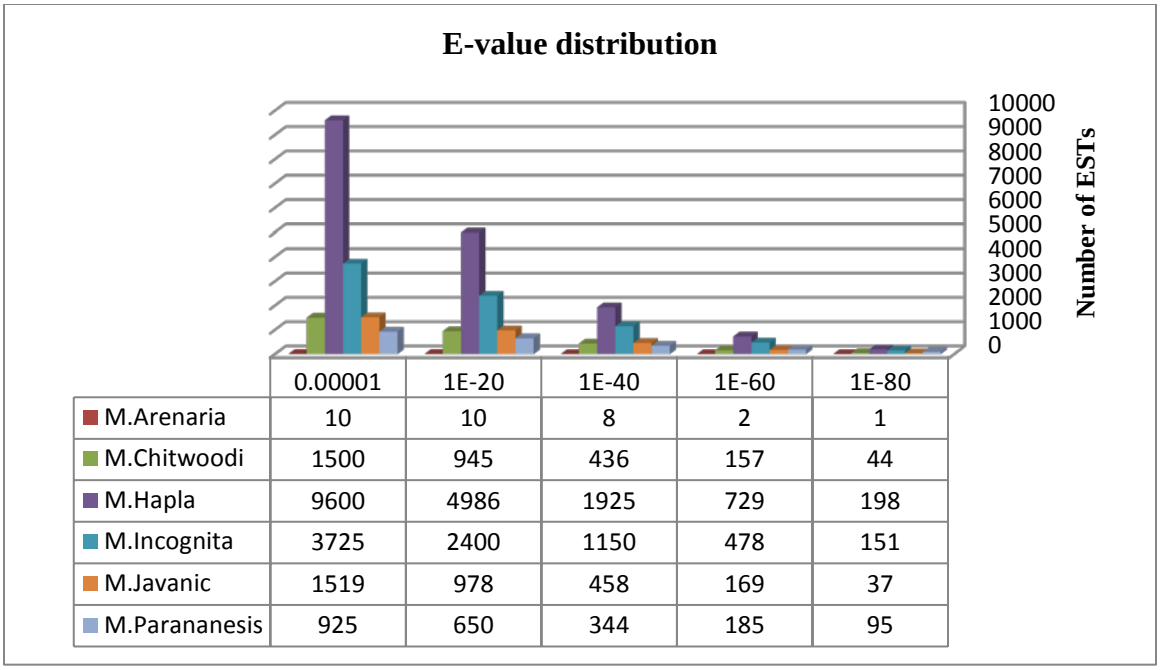
2.3.1 Pairwise alignment

We performed pairwise sequence alignments of the six selected *Meloidogyne.spp* unigene sequences versus the *C. elegans* sequences. Overall, we identified 7 candidate genes between *M. arenaria* and *C. elegans* with 2 lethal phenotypes; six are lethal at the embryonic stage while the one is lethal at the larval stage and one is lethal at both embryonic and larval stages. We identified 1491 candidate genes between *M. chitiwoodi* and *C. elegans* with 6 lethal phenotypes; lethal at the embryonic stage, lethal at the late embryonic stage, lethal at the larval stage, lethal at the early larval stage, lethal at the adult stage, and lethal in all stages of development. We identified 2934 candidate genes between *M. hapla* and *C. elegans* with 8 lethal phenotypes; lethal at the embryonic stage, lethal at the late embryonic stage, lethal at the early embryonic stage, lethal at the larval stage, lethal at the late larval stage, lethal at the early larval stage, lethal at the adult stage, and lethal in all stages of development. We identified 3229 candidate genes between *M. incognita* and *C.*

elegans with 6 lethal phenotypes; lethal at the embryonic stage, lethal at the late embryonic stage, lethal at the larval stage, lethal at the early larval stage, lethal at the adult stage, and lethal in all stages of development. We identified 1962 candidate genes between *M. javanica* and *C. elegans* with 8 lethal phenotypes; lethal at the embryonic stage, lethal at the late embryonic stage, lethal at the early embryonic stage, lethal at the larval stage, lethal at the late larval stage, lethal at the early larval stage, lethal at the adult stage, and lethal in all stages of development. Finally, we identified 779 candidate genes between *M. paranaensis* and *C. elegans* with 6 lethal phenotypes; lethal at the embryonic stage, lethal at the late embryonic stage, lethal at the larval stage, lethal at the early larval stage, lethal at the adult stage, and lethal in all stages of development.

2.3.2 Significance groups

The genes are divided according to our focus and efforts into five E-value cut-offs for all six species (see figure 4), E-values between 0 and 1E-80, > 1E-05 (Group1), >1E-20 (Group2), >1E-40 (Group3), >1E-60 (Group4), >1E-80 (Group5). We divided our sequence comparison results into five groups to assess whether a given alignment constitutes evidence for homology; it helps to know how strong an alignment can be expected from chance alone. Chance can mean real but non-homologous sequences, real sequences, or sequences that are generated randomly based upon a DNA or protein sequence model. A $1E{-}5$ is the standard of an alignment that means alignment is highly unique, and not due to error.



	0.00001	1E-20	1E-40	1E-60	1E-80
M.Arenaria	10	10	8	2	1
M.Chitwoodi	1500	945	436	157	44
M.Hapla	9600	4986	1925	729	198
M.Incognita	3725	2400	1150	478	151
M.Javanic	1519	978	458	169	37
M.Parananesis	925	650	344	185	95

Figure 4: The chart is depicting the number of ESTs (lethal and not lethal) in *Meloidogyne.spp* that showed similar identity with our local *C.elegans* protein database by using the BLASTX algorithm. The standard e-value 1*E*-5 was used but in our experiment we were able to identify higher similarity between RKN and C. elegans. The number of genes in the table part in figure 1 and particularly to the last column 1*E*-80 represent the most significant similarity between the unknown (RKN) and the well-known (C. elegans).

2.3.3 Lethal Stages

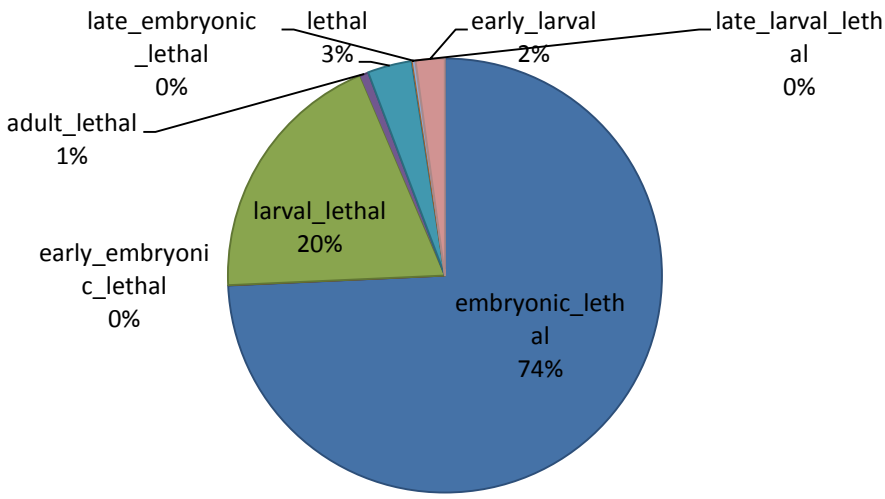


Figure 5: Distribution of Lethal Genes across Stages

We were able to identify conserved genes that potentially were lethal at all stages of Meloidogyne.spp development and they were highly identical to C.elegans using 1*E*-5 as the cut-off E-value. Of these, (74%) were identified as producing lethal phenotype in C.elegans if mutated or silenced at the embryonic stage, (20%) were identified as producing lethal phenotype at the larval stage, (3%) were identified as producing lethal phenotype at all stages, (2%) were identified as producing lethal phenotype at the early larval stage, (0.37%) were identified as producing lethal phenotype at the adult stage, (0.1%) were identified as producing lethal phenotype at the late embryonic stage, (0.03%) were identified as producing lethal phenotype at the late larval stage, and finally we identified two gene that produces lethal phenotype at the early embryonic stage. These genes were distributed over the six species of

Meloidogyne.spp as followed in (Table2).

Stage of Development RKN Species	Embryonic	Early Embryonic	Larval	Adult Lethal	Lethal	Early Larval	Late Larval	Late Embryonic
M. arenaria	6	0	1	0	0	0	0	0
M. chitwoodi	1105	0	300	12	44	29	0	1
M. Hapla	2145	1	539	12	147	85	2	3
M. Incognita	2319	0	794	16	51	40	0	9
M. Javanica	1590	1	241	26	54	37	5	8
M. Paranaensis	560	0	134	2	47	35	0	1

Table 2 is depicting the number of genes that are lethal at different stages of *Meloidogyne.spp* development.

2.4 Conclusion

Chemical control of the RKN is not used because of the economic and environmental prohibition. Therefore we decided to control RKN biologically. The application of RNA interference to genomic studies of C. elegans has been very successful in helping to identify genes that play an important role in the life cycle of RKN. Our analysis yielded many candidate lethal genes between the major six species of RKN and C. elegans. Some of these genes have already been tested and they found to be lethal to the RKN. Not only we were able to discover the candidate lethal genes in RKN but also we were able to tell exactly which genes are lethal at which stage of the RKN life cycle development.

Chapter 3

Microarray Data Analysis

3.1 Introduction on Microarray

Microarray is a 2D array that is used for monitoring the level of expression of large number (thousands) of genes simultaneously in a single experiment. Microarray is a method to figure out which genes are likely to be used in a particular tissue at a particular time under a particular set of conditions. The output of a microarray experiment is called a gene expression profile (Gibson 2003). Different types of arrays are available but the main goal is to determine whether gene expression profiles differ among samples from different classes and identify which genes are differentially expressed between classes. Recently, most of the microarray analysis used to look at the plant defense responses in the existence of abiotic stresses such as drought, cold, salt, and temperature (Kawasaki *et al.* 2001). RKN can be easily recognized by the "knots" or "galls" that form where they feed on roots. RKN can cause severe morphological and physiological changes in plant cells (Caillaud *et al.* 2008). RKN damage to soybean (*Glycine max*) can be severe, especially when fields previously planted in cotton are rotated into soybean (Bennett *et al.* 2010). Within the nematode's infection stage, different proteins are produced to help the nematode establish a feeding site. Some of the proteins secreted by the nematode are injected into host cells. Other proteins secreted by the nematode may interact with the host's extracellular receptors (Ibrahim *et al.* 2011). Recent targets to control nematode could be developed via the identification of genes that are involved in the establishment of

the nematode in the host plant particularly in the formation of the permanent feeding site for the nematode (Ibrahim *et al.* 2010)

Similarly, gene expression is altered roots and in syncytial cells in soybean roots infected by SCN (Klink *et al*. 2007). They used microarrays to study gene expression in Laser Capture Microdissected (LCM) syncytium cells to monitor sensitive and resistant reactions of soybean during infection with SCN. Some of the genes were shown to be up- regulated (induced) while the others were shown to be down-regulated (suppressed) in both sensitive and resistant responses.

3.2 Introduction to SCN

3.2.1 Damage

SCN is the dominant pathogen of soybean (*Glycine max*) and it is responsible for an estimated $0.46–$0.82 billion in production losses, annually, in the U.S. and about $15 billion worldwide (Wrather and Koenning 2006). SCN infection of soybean causes various symptoms that may include chlorosis, root necrosis, loss in seed yield and suppression of root and shoot growth. In general, the infection of plants by parasitic nematodes accounts for losses of approximately $157 billion U.S. dollars worldwide, annually (Abad *et al*. 2008). Chemical control (soil fumigation) of SCN is not normally used because the economic and environmental costs are prohibitive. The use of naturally occurring genes to control SCN and other pests is generally more acceptable ecologically and economically than chemical control and is the easiest choice for SCN control.

3.2.2 Life Cycle

SCN is a sedentary root endoparasite; these nematodes invade the root and partially recognize root cell function to satisfy their nutritional demands for development and reproduction. The *H. glycines* life cycle is approximately 1 month in duration (Jung and Wyss 1999). The eggs within cysts can lie dormant in the soil for years. The eggs contain the first stage juveniles (J1). Eggs hatch as pre-infective second stage juveniles (pi-J2) that migrate toward the root and become infective second stage juveniles (J2). The J2's are equipped with a robust stylet to pierce neighboring root cells to facilitate intracellular migration towards the differentiating vascular cylinder, initiating the feeding site. Those feeding site initials then fuse with neighboring cells producing a syncytium, a structure that contains approximately 200 merged root cells, each with a greatly enlarged nucleus and metabolically highly active cytoplasm (Jung and Wyss 1999). Syncytia associated with the male nematodes are less extensive than those associated with female nematodes, as males feed only until the end of the third juvenile stage (J3). Then the males become vermiform again while they molt to the J4 stage, and after the last molt they emerge through the juvenile cuticles in search of females ready for copulation (see figure 6).

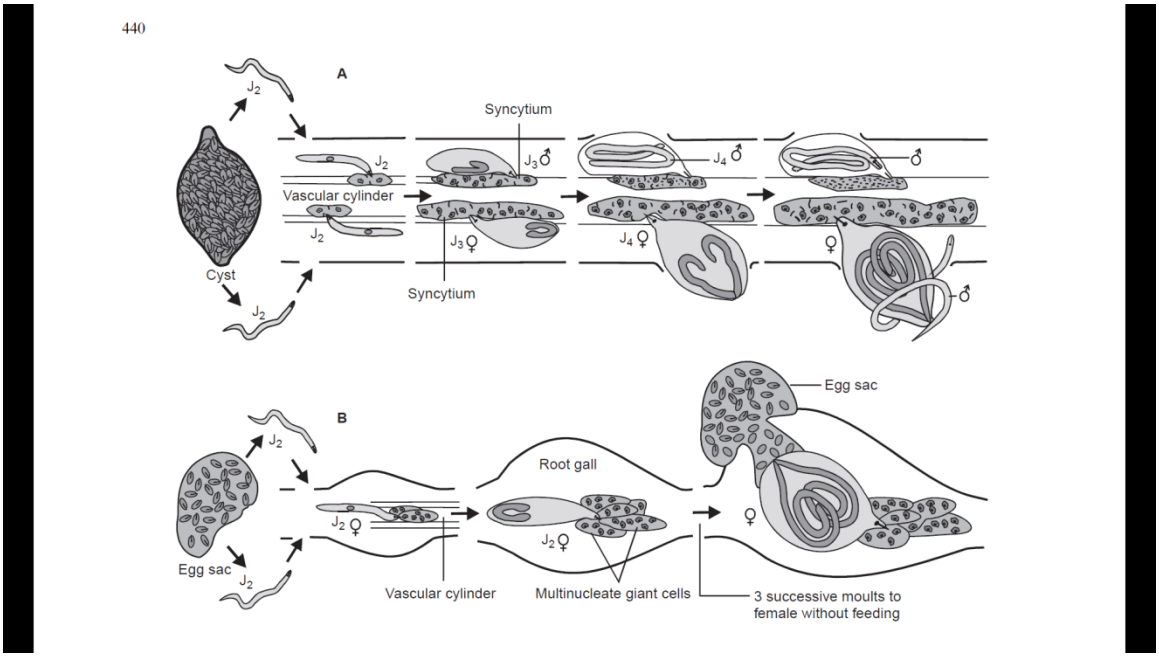


Figure 6: Life cycle of SCN. From (Jung and Wyss 1999).

At the ultra-structural level, syncytia exhibit the characteristic feature of high metabolic activity, as revealed by hypertrophied nuclei and a remarkable increase in cytoplasmic density accompanied by a proliferation of cell organelles involved in synthesis transport and energy supply (Jung and Wyss 1999). Wall ingrowths, characteristic of transfer cells, are formed where these feeding structures contact xylem vessels. The function of these ingrowths is for efficient food withdrawal (Grundler and Bockenhoff 1997). The molecular triggers involved in feeding site induction are not yet known. However, it is generally believed that secretary proteins released through the nematode stylet orifice play a decisive role in the induction process.

3.2.3 Soybean resistance to SCN

Investigators have identified several genes that respond to nematode invasion, including extensin (Eycken *et al*. 1996), catalase, cyclin (Niebel *et al*. 1995; Vaghchhipawala *et al*. 2001), β-1,3 endoglucanase (Vaghchhipawala *et al*. 2001), adolase and GTP-binding protein (Hermsmeier *et al*. 1998). The complexity of the soybean –SCN interaction suggests that there are even more genes involved in the soybean response to nematode infection. There are several known loci in soybean encoding resistance to SCN but the contributions of different loci to the resistance response at the molecular level are not known. Furthermore, the functions of genes involved in resistance are unknown. The nematode penetrates susceptible and resistant cultivars to the same extent, but resistant cultivars prevent most of the attacking females from reaching maturity and producing viable eggs. Extensive microscopic studies indicate that from 18 hours through the first two to three days, smooth and rough endoplasmic reticulum increase and cell wall dissolve in syncytia of resistant and susceptible cultivars (Endo 1964; 1965; 1991; Kim *et al*. 1987; Mahalingam and Skorupska 1996). However, at five through nine days after inoculation, the response of resistant and susceptible soybean cultivars are different. The syncytia continue to develop and expand in susceptible soybean roots, but resistant soybean roots show extensive cell wall depositions; the cells are almost devoid of organelles, and there is syncytium degeneration and cell necrosis. To gain more details about molecular mechanisms underlying the basis of SCN resistance, global gene expression patterns in SCN-resistant and susceptible soybean cultivars after exposure to the nematodes need to be examined (klink *et al*. 2009).

3.3 Experimental Design

Experimental design is an important step when planning for a microarray experiment. A good design will allow making the most of your resources. In the soybean infected with root knot, three time points [Control, 12 days post infection (dpi), and 10 weeks post infection (wpi)] were studied. A standard reference design was used for these microarray experiments. The reference or control was RNA that was extracted from soybean roots not infected with *Meloidogyne Incognita*. The 12 dpi time point had been chosen as the small galls are formed in soybean by *M. incognita*. While the 10 wpi was chosen as the large galls are formed. Therefore, these time points were interesting to study the plant genes expressions in plant roots after infection. In our experiment, three replicates of each time point were used to maximize the accuracy of our results. The objective from this study was to identify possible gene targets for manipulation to develop resistant plants to RKN by using gene silencing technology. Microarray gene expression analyses were conducted according to (Klink *et al*. 2007; 2010) using the GeneChip Soybean Genome Array (Affymetrix_; Santa Clara, CA). The soybean GeneChip_ contains 37,500 *G. max* probe sets. Details of the GeneChip_ soybean genome array are available at the Affymetrix_website: (http://www.affymetrix.com/support/technical/datasheets/soybean_datasheet.pdf).

The microarrays were hybridized and scanned at the Laboratory of Molecular Technology, SAIC-Frederick, National Cancer Institute at Frederick, Fredrick, MD, USA. Also PAICE (Pathway Analysis and Integrated Coloring of Experiments) software was used to visualize genes expressions and their biochemical pathways (Hosseini *et al*. 2012). Numbers of different pathway genes were identified to be

affected by nematode infection. This study was to provide information of the interaction between *Meloidogyne incognita* and soybean to identify possible lethal gene targets to develop wide resistance of plants to RKN by using gene silencing technology or to over express certain soybean genes.

3.4 Materials and Methods

3.4.1 Plant and nematode materials

Both *Glycine max* cv Williams 82 and *M. incognita* were grown in a greenhouse at the United State Department of Agriculture Soybean Genomics and Improvement Laboratory, Beltsville, MD, USA. Soybean seeds were surface sterilized, and then germinated on water agar plates for 3 days at room temperature in the dark. Soybean seedlings were grown in Promix (Premier Horticulture INC., Quakertown, PA, USA) for one week in 20 x 20 x 10 cm flats, then moved to sand (The Stone Store, Hanover, MD). *M. incognita* eggs were harvested from roots of *G. max* cv Williams 82 at 2-4 months after the initial inoculation in sterile water at room temperature as described in (Meyer *et al.* 2000). Eggs were incubated in sterile water at room temperature to promote hatching. After 2 days the juveniles were collected and concentrated by centrifugation to approximately 3,000 J2/ml. 7 day old soybean seedlings (cv. Williams 82) roots were used to inoculate three thousands *M. Incognita* eggs J2s in sterile water. The control replicates received the same volume of sterile water. After 12 dpi, 10 wpi, of exposure to the nematodes, the roots of both control and infected roots were washed gently in flowing tap water to remove any debris and J2s that had not infected the roots. The infected roots were stained using a modified protocol of

(Bybd *et al.* 1983) and (Mahalingam *et al*. 1996) cut to 2 cm segments, and placed in a small beaker, then soaked in 20-30 ml of 10 % commercial Clorox (chlorine bleach, 5.25 % NaOC1) for 3 min. Then roots were washed in tap water and then transferred into a glass bottle containing 20 ml of distilled water and left to boil in a microwave with loosened caps. Then roots were transferred into a glass bottle containing 500 µl of acid fuschin (0.15g/10ml H2O) and 500 µl of glacial acetic acid and left to boil in a microwave twice. Then roots were left at room temperature to cool down before removing the extra unnecessary stain with running tap water using Mira cloth on the top of the bottle. Then the roots were left to distain for two hours to overnight in A 20 ml of clearing reagent (1/3 lactic acid+ 1/3 glycerol + 1/3 distilled water). The roots were rinsed in deionized water and flash frozen in liquid nitrogen at -80º C until use. Figure 7 is showing the soybean roots after infection by *M. incognita* at different time points (12 dpi and 10 wpi).

The frozen root tissue was ground to a fine powder using mortar pestle chilled in liquid nitrogen. Total RNA was extracted from 100 mg each of the three different root samples (12 dpi, 10 wpi, and control uninfected plants) using The Ultra Clean Plant RNA Isolation Kit (MOBIO, Carlsbad, CA).

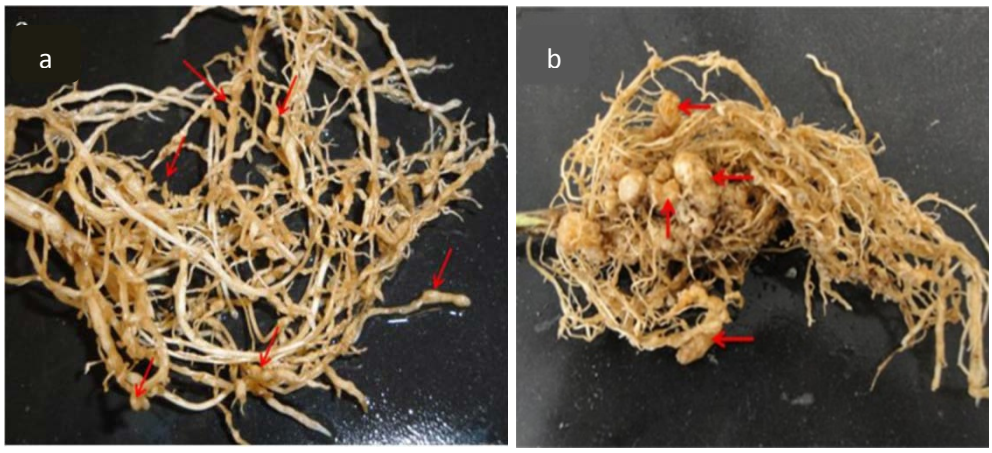


Figure 7: Soybean roots infected by M. incognita at different time points.

(a) galls formed on soybean roots at 12 dai; red arrows point to the galls; (b) galls formed on soybean roots at 10 wai; red arrows point to the galls (Ibrahim *et al.* 2011).

3.4.2 RKN microarray data processing

In the Affymetrix microarray platforms, gene expression is measured using probe sets consisting of 11 to 20 perfect match (PM) probes (25 nucleotides in length) complementary to target mRNA sequences. Each probe set also has the same number of mismatch (MM) probes, in which the nucleotide number 13 has been changed to its complement. The PM probes are designed for gene specific hybridization while the control MM probe measurements are designed to comprise most of the background non-specific binding. The measured probe intensities and locations from a hybridized microarray are stored in a CEL file while the information for each probe set such as probe set IDs, and locations on the array is stored in a CDF library file. The probe sequence information is stored in a sequence file (FASTA or tab-separated format). MATLAB version (R2010b) Bioinformatics toolbox was used to preprocess

Affymetrix probe-level expression data based on Robust Multi-array Average (RMA) procedure in three steps: background adjustment, normalization, and summarization at the probe set level as a measure of the expression level of corresponding mRNA.

Importing Affymetrix Data

The RKN experiment was conducted using GeneChip Soybean Genome Affymetrix array, and the probe level data provided by Affymetrix were stored in what we call the CEL files, these usually have extension .CEL. Each CEL file is storing information for one replicate. According to our experiment we have three sample points (Uninfected, 12 dpi, and 10 wpi) and three replicates for each sample point. Therefore, we had nine CEL files in our experiment. The function AffyRead had been used to import these data into Matlab workspace. Information related to each probe is contained in the Affymetrix Soybean CDF library file. One easy way of reading in the CEL files is to have all CEL files in one directory and set the working directory to that directory. To set the working directory you can use the menu commands on either Windows or OS X. Then you can simply use the AffyRead function, as shown in the code below.

Cel1 = affyread (MacDonald_Heba_01_102308.CEL);

The provided file, rknsamples.txt, contains a list of CEL filenames used for this experiment, to which replicate they belong. The file was loaded into MATLAB workspace using the code below;

[replicates] = textread('rknsamples.txt','%q%q');

The data was loaded from the CEL files into a PM probe intensity structure from the directory where the CEL files are stored, and pass in the path to where you store the library files.

libPath = 'C:\primatedemo\libfiles';

pmStruct = celintensityread(replicates, 'cdfname', 'cdfpath', libPath)

To check the number of CEL files loaded we used the next code

nSamples = pmStruct.NumChips

The matrix of PM intensities from multiple CEL files is stored in the 'PMIntensities' field. The matrix column indices correspond to the order of the CEL files read, and each row corresponds to a probe

pmMatrix = pmStruct.PMIntensities;

For accuracy we determined the number of probe sets on the Soybean GeneChip.

nProbeSets = pmStruct.NumProbeSets;

Preprocessing Probe-Level Expression Data

In general, preprocessing Affymetrix probe-level expression data consists of these three steps: background adjustment, normalization, and summarization. You can use the methods based on the RMA procedure to correct the background, normalize the data and calculate the expression values.

First, we used the RMA background adjustment method to correct PM intensities and the estimated background adjustment of soybean chip using the code:

pmMatrix_bg = rmabackadj(pmMatrix);

Note that each row in pmMatrix corresponds to a perfect match (PM) probe and each column in pmMatrix corresponds to an Affymetrix CEL file.

Then, normalization was applied using the RMA procedure to normalize the probe-level data with a quantile normalization method called **quantilenorm** to normalize the background adjusted PM intensities in the RKN data, after that we used the boxplot to compare normalized intensity distributions across several chips (figure 8).

NormMatrix = quantilenorm(BackgroundAdjustedMatrix);

maboxplot(log2(pmMatrix_bgnorm(:, 1:9)), samples(1:9), 'title','Normalized Intensities')

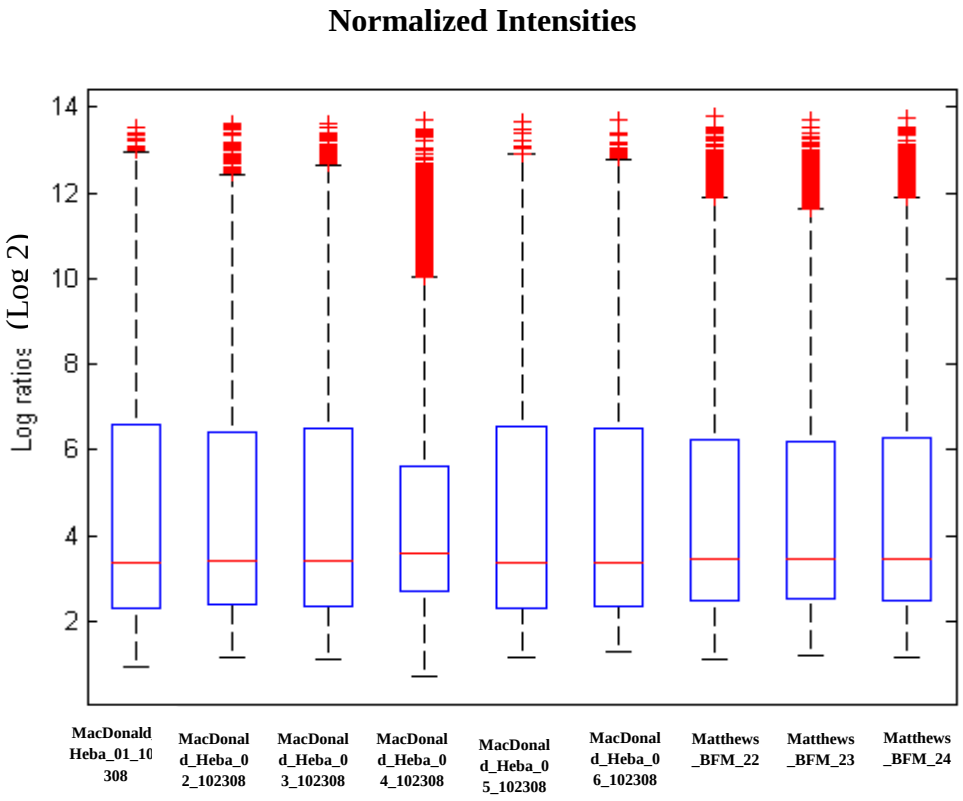


Figure 8: Boxplot of Soybean roots infected by M. incognita at different time points.

Finally, a median polish procedure is applied to the PM intensities in summarization. To calculate the expression values, use **rmasummary** to summarize probe intensities of each probe set. The expression values are the probe set intensity summaries on a log-2 scale.

ExpressionMatrix = rmasummary(probeIndices, NormMatrix);

The convention for probe indices is, for each probe set, to label each probe 0 to N – 1, where N is the number of probes available in the probe set.

To remove the housekeeping genes and genes with empty gene symbols we used strmatch. Affymetrix housekeeping genes IDs have prefix "AFFX-"

indices = strmatch('AFFX', pmStruct.ProbeSetIDs);
primateGenes(indices) = [];
primateValues(indices, :) = [];

Many of the genes on the chip may have only small variability across the samples. These genes should be removed by non-specific filtering. We filtered out those genes with very low absolute expression values with an intensity filter.

mask = genevarfilter(ExpressionMatrix);
ExpressionMatrix = ExpressionMatrix(mask, :);
FilterProbeSetIDs = ProbeSetIDs(mask);

3.4.3 SCN microarray data processing

The SCN experiment was conducted using GeneChip (Cat. #900526; Affymetrix_; Santa Clara, CA) Soybean Genome Affymetrix array according to (Klink *et al*. 2007). The experiment was conducted using four time points [0 hours post infection (hpi), 12 hpi, 3 days post infection (dpi) and 8 dpi] were studied and two replicates for each sample point. According to our objective of studying the changes of the gene expression profile for the soybean root after infection with RKN and SCN and compare it to each other we decided to study and analyze only two of those four samples [0 (hpi) and 8 dpi].

Similarly, MATLAB version (R2010b) Bioinformatics toolbox was used to preprocessing Affymetrix probe-level expression data based on Robust Multi-array Average (RMA) and the probe level data provided by Affymetrix were stored in the CEL files, one for each replicate. We preprocessed Affymetrix probe-level expression data following same three steps: background adjustment, normalization, and summarization. We used the methods based on the RMA procedure to correct the background, normalize the data and calculate the expression values.

First, we used the RMA background adjustment method to correct PM intensities and the estimated background adjustment of soybean chip. Then, we normalized the probe-level data with a quantile normalization using the **quantilenorm** function to normalize the background adjusted PM intensities in the SCN data. We used the boxplot function in Matlab to compare normalized intensity distributions across several chips (see figure 9).

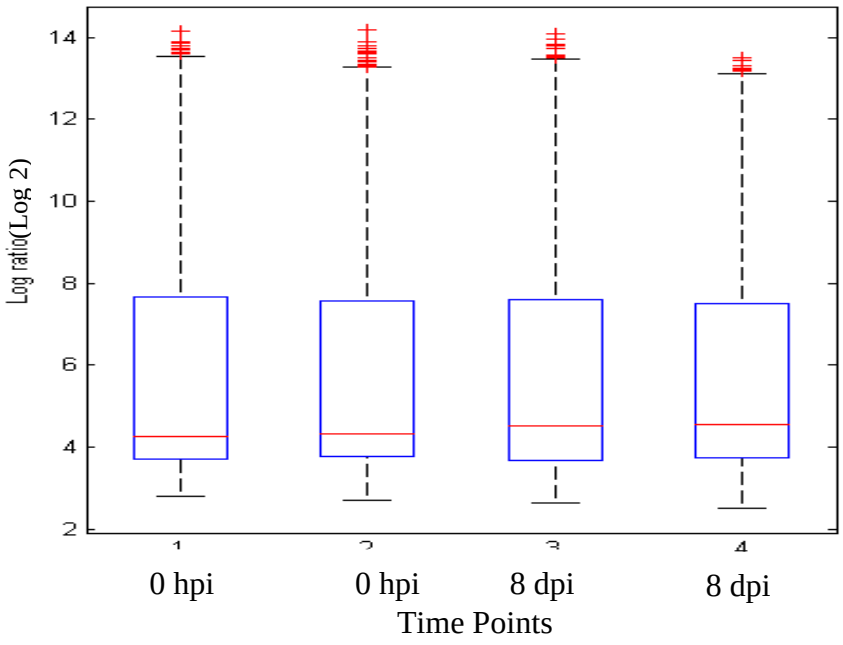


Figure 9: Boxplot of Soybean roots infected by *H. glycine* at different time points

Finally, we used the **rmasummary** function to summarize probe intensities of each probe set and to calculate the expression values. The expression values are the probe set intensity summaries on a log-2 scale. All the genes on the chip that have only small variability across the samples were removed by non-specific filtering using **genevarfilter** function.

3.4.4 Annotation

Blast2Go tool was used to associate individual sequences and related expression information with biological function. B2Go is an effective approach for automated functional annotation. The benefit of using functional annotation is to categorize large amounts of genes into functional classes, which can be very useful to understand the physiological meaning of them. B2G has been used to allow automatic and high

sequence annotation and integrate functionality for annotation-based data mining and finally assigns GO terms to the query sequence.

First step was to obtain GO terms to find sequences similar to a query set by Blast searching. The search was done on the public database NCBI nr using blastX algorithm. The statistical significance threshold for reporting matches against database sequences was 1.0E-3. After that the mapping was made In order to retrieve GO terms associated with the obtained hits. Using blast hit gene identifiers (gi) and gene accessions retrieves all GO annotations for the hit sequences, together with their evidence codes (EC). By the end of the mapping processes, we identified a set of candidate annotations from different hits of different similarity levels for each query sequence. Then InterProScan was made to combine different protein signatures from different databases into a single searchable resource. These signatures mainly represent proteins belonging to the same family, functional domains, or active sites and so we were able to specify the potential function of our protein. Many protein signatures have been integrated into the InterPro database and therefore have GO terms associated with them. In this way, we used InterProScan to associate functional information and GO terms to our protein of interest quickly.

3.5 Microarray results

3.5.1 Exploring Differential Gene Expression in RKN

The next step is to examine the expression values for genes that are differentially expressed between the sample points. We used t-test to identify genes that are differentially expressed between the two groups (12dpi and 10wpi) against the control

or uninfected group so we can get an impression of differential gene expression by looking at the histograms of t-test results (figure 10a&b). In order to do the t-test we calculated the average of the gene expressions from all three replicates for each time point. The p-value cutoff was 0.05 to identify genes that are considered to have statistical significance.

[pvalues, tscores] = mattest(Average12dpiData, ControlData, 'Showhist', true', 'Showplot', true);

[pvalues, tscores] = mattest(Average10wpiData, ControlData, 'Showhist', true', 'Showplot', true);

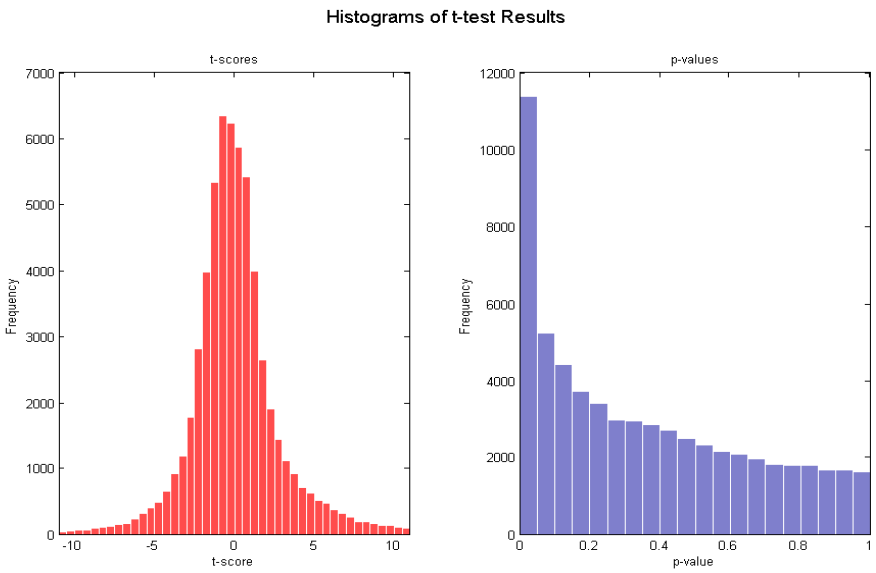


Figure 10.a: histogram of t-test result to identify genes that are differentially expressed between the 12dpi and control.

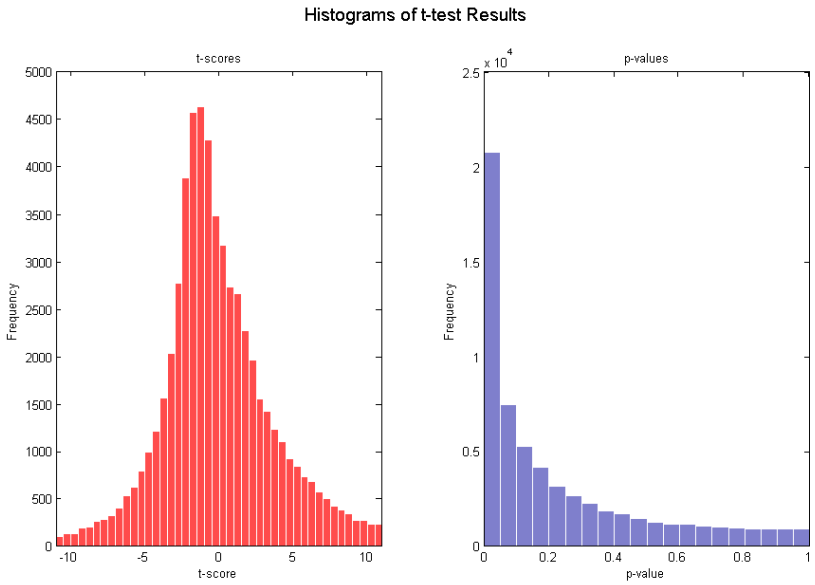


Figure 10.b: histogram of t-test result to identify genes that are differentially expressed between the 10 wpi and control.

A gene is considered to be differentially expressed between the two samples if it shows both statistical significance and biological significance. The biological effect is the log2 of the expression ratio between two groups. Then, we plot the –log10 of p-values against the biological effect in a volcano plot. We produced Volcano plots that can visualize gene expression fold-change based on significance versus fold-change on the y- and x-axes, respectively. The plots were produced using fold-changes >= 1.5 and p-values <=0.05 with respect to the control or untreated roots (figures 11.a&b).

Volcano plot is a type of scatter-plot that is used to identify changes in gene expression. It plots significance versus fold-change on the y- and x-axes, respectively. The y-axis is the log 10 of the p-value. The x-axis is the log 2 of the fold

change between the two conditions (up and down) appear equidistant from the center. Plotting points in this way results in two regions of interest in the plot: those points that are found towards the top of the plot that are far to either the left- or the right-hand side (the darker points) and the points found in the middle and towards the bottom (the lighter points). The darker points to the right represent all the induced genes while the ones to the lift represent all the suppressed genes. The lighter points in the middle and at the bottom represent all the genes that are not differentially expressed.

diffStruct = mavolcanoplot (Average12dpiData, ControlData, pvalues, 'labels', primateGenes)

diffStruct = mavolcanoplot (Average10wpiData, ControlData, pvalues, 'labels', primateGenes)

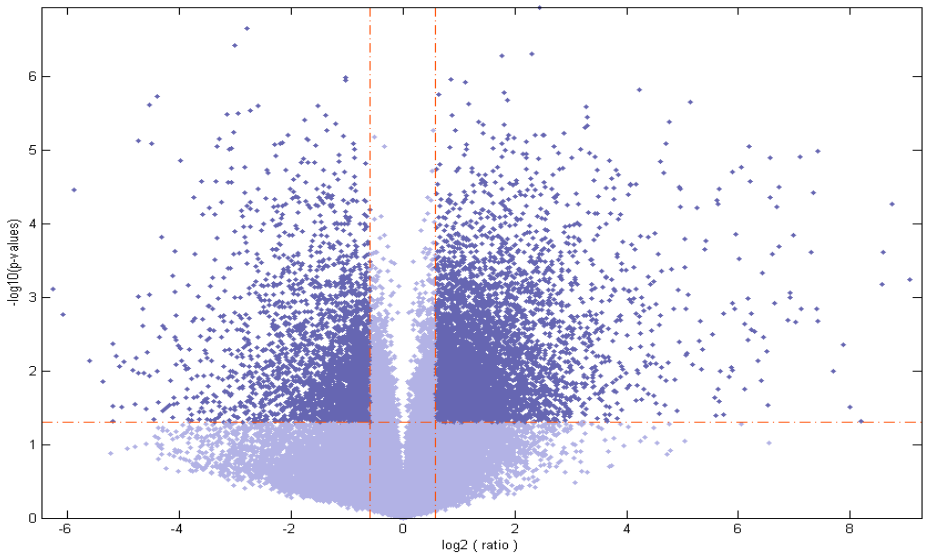


Figure 11.a: Volcano plot between 12dpi and Untreated (RKN).

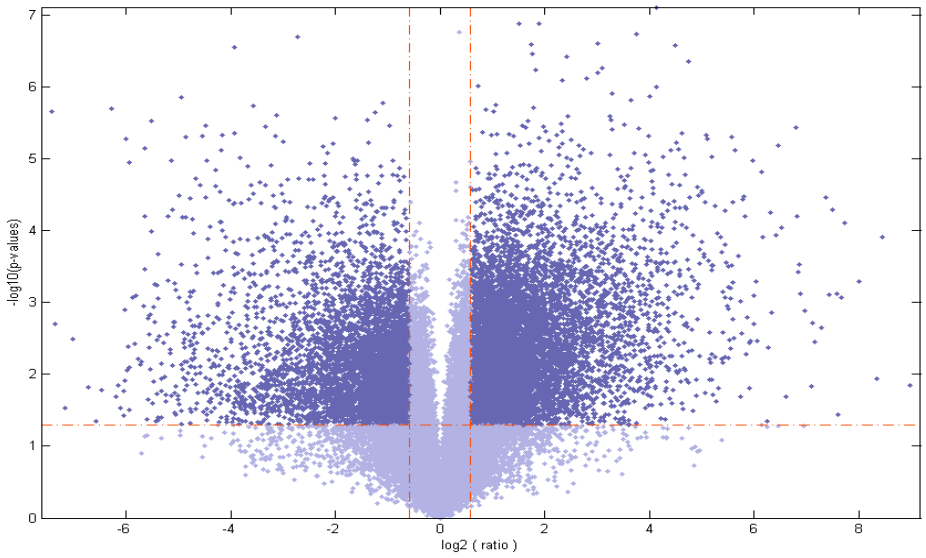


Figure 11.b: Volcano plot between 10wpi and Untreated (RKN).

3.5.2 Exploring Differential Gene Expression in SCN

Similarly, we examined the expression values for genes that are differentially expressed between the sample points. We used the t-test to identify genes that are differentially expressed between the one group (8dpi) against the control or uninfected group so we can get an impression of differential gene expression by looking at the normal quantile plot of t-scores (see figure 12) and the histograms of the t-test results providing a graphical view of how different the gene expression distribution is between the two sample points (see figure 13). In order to do the t-test, we calculated the average of the gene expression from all three replicates for each time point. The p-value cutoff was 0.05 to identify genes that are considered to have statistical significance.

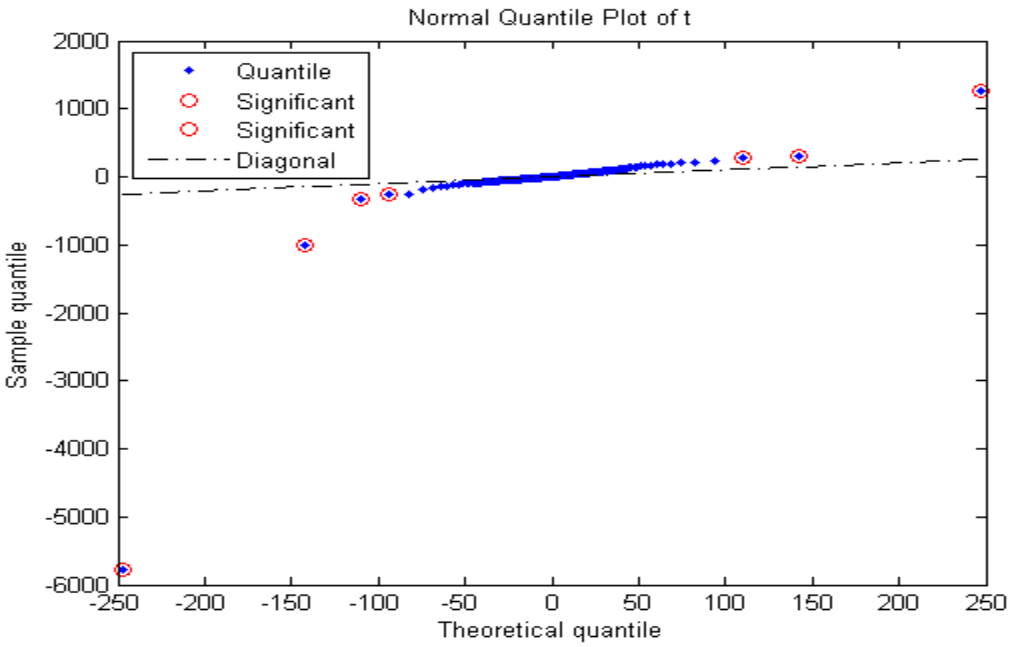


Figure 12: normal quantile plot of t-scores showing the gene expression distribution of the two sample points (8dpi and untreated).

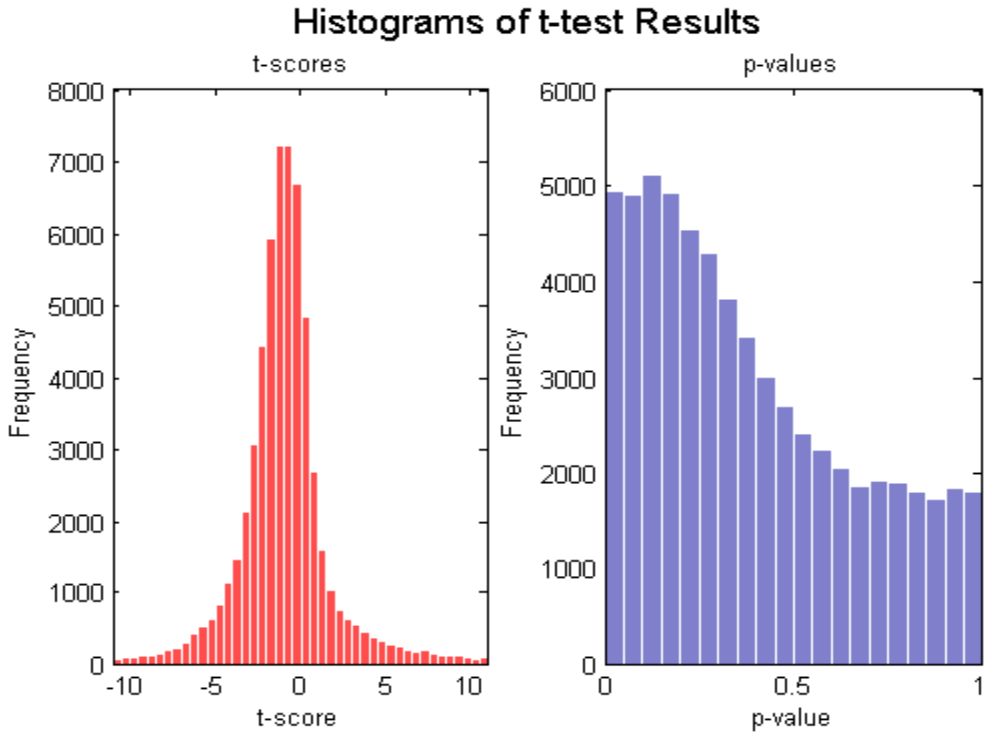


Figure 13: histogram of t-test result to identify genes that are differentially expressed between the 8dpi and control.

We produced Volcano plots that can visualize gene expression fold-change based on significance versus fold-change on the y- and x-axes, respectively. The plots were produced using fold-changes >= 1.5 and p-values <=0.05 with respect to the control or untreated roots (see figure 14).

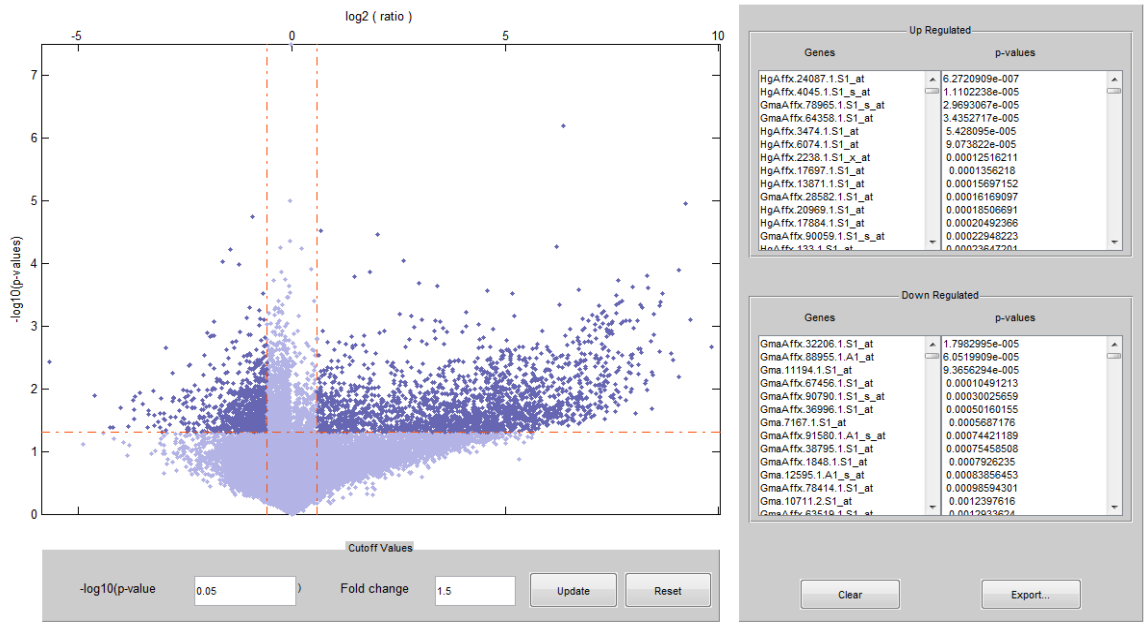


Figure 14: Volcano plot between 8dpi and Untreated (SCN).

3.5.3 Soybean infection by SCN vs. RKN

In our comparison we focused on one time point of each experiment. SCN Infection during the first 8dpi was focused on for this analysis because syncytial cells complete the incompatible reaction by 8dpi under the experimental condition in *G.max*. In the RKN infection 12dpi time point was chosen to contrast active nematode feeding and small galls formation with plant gene expression in mature infection. At 12dpi, we identified a total of 7,494 transcripts as being differentially expressed. 5,221 of those transcripts were identified to be induced while only 2,273 transcripts were identified to be down regulated. Transcripts encoding nodulin (BE020011) increased the most at ~539-fold (see table 3). The most down regulated gene was (BF070134) with homology to TR: O48832 putative physical impedence protein; its transcripts were 77-fold lower than in the control (see table 4).

Table 3: List of the 25 most induced genes in soybean (cv.Williams 82) at 12 dpi by *M. incognita* (P-value≤0.05) and (Fold Change>1.5)

Transcript ID	*P*-Value	Fold Change	Public ID	Gene Title
GmaAffx.77823.1.A1_at	0.00057157	538.8493	BE020011	nodulin-16-like
GmaAffx.4508.1.S1_at	5.32E-05	432.148	BI316036	---
GmaAffx.2446.1.S1_at	0.00024083	385.7596	AI973819	leghemoglobin C1-like
GmaAffx.39680.1.S1_at	0.00065947	380.4562	BI700248	---
Gma.4132.2.S1_s_at	0.00142576	169.9623	BI700028	uncharacterized LOC100779526 /// uncharacterized LOC100783288 /// uncharacterized LOC100785417 /// uncharacterized LOC100803066 /// uncharacterized LOC100805736 /// uncharacterized LOC100807853 /// uncharacterized LOC100811611 /// uncharacterized LOC10081
GmaAffx.68724.1.A1_at	0.0002422	156.9751	BI970438	Cytokinin dehydrogenase 3-like
GmaAffx.88845.1.S1_at	0.00143993	138.5036	CK605838	basic form of pathogenesis-related protein 1-like
Gma.7.1.S1_at	1.21E-05	138.2888	X96792.1	uncharacterized LOC100798802 /// nodulin
Gma.27.1.S1_at	0.0021704	130.8192	X69156.1	---
Gma.16831.1.S1_at	0.00014377	126.6803	M16976.1	N-75 protein
Gma.8768.1.A1_at	0.0008766	121.4918	BU764586	MATE efflux family protein FRD3-like
Gma.16709.1.S1_s_at	0.00101319	121.009	BQ742929	isoflavone 2'-hydroxylase-like
GmaAffx.84446.1.S1_at	0.00203414	118.3307	AI988577	uncharacterized LOC100527392
Gma.5121.1.A1_at	0.00020086	105.8646	AW309540	subtilisin-like protease SDD1-like
GmaAffx.59606.1.S1_at	5.91E-05	102.2874	BI787321	ser/Thr-rich protein T10 in DGCR region-like
Gma.13.1.S1_at	0.00118337	100.4082	D13506.1	early nodulin
GmaAffx.36484.1.S1_s_at	0.00025743	97.56286	BI425441	basic form of pathogenesis-related protein 1-like
Gma.4829.1.S1_at	1.25E-05	95.17569	AW309606	peroxidase 4-like
Gma.22.1.S1_at	4.34E-05	94.4046	X59903.1	nodulin 24
Gma.19.1.S1_at	0.0289147	91.8364	X05091.1	nodulin-22
GmaAffx.84435.1.S1_s_at	0.00344773	87.53476	AI973597	nodulin-20
Gma.14.1.S1_at	0.01347984	85.24384	D13502.1	early nodulin
Gma.265.1.S1_at	0.0072604	81.4847	X03979.1	nodulin (E27) (aa 1-360)
GmaAffx.42847.1.S1_s_at	0.00289687	78.23046	BU578388	basic form of pathogenesis-related protein 1-like
Gma.4112.1.S1_at	0.00277985	74.88904	BI700392	miraculin-like

Table 4: List of the 25 most suppressed genes in soybean (cv.Williams 82) at 12 dpi by *M.incognita* (P-value≤0.05) and (Fold Change<1.5)

Transcript ID	*P*-Value	Fold Change	Public ID	Gene Title
GmaAffx.63207.1.S1_at	0.000767	-77.1037	BF070134	uncharacterized LOC100802385
Gma.6552.1.S1_at	0.001713	-67.5555	BQ627716	metal transporter Nramp5-like
GmaAffx.53768.1.S1_at	3.42E-05	-59.3	BG789629	allene oxide synthase
Gma.5787.2.S1_at	0.007171	-48.7862	CA784580	GDP-mannose 4,6 dehydratase 1-like
soybean_rRNA_1479_at	0.014013	-41.2084	soybean_rRNA_1479	---
GmaAffx.49392.1.S1_at	0.029989	-36.7115	BE805053	uncharacterized LOC100784458
GmaAffx.90411.1.S1_s_at	0.004218	-36.6703	CF806136	GDP-mannose 4,6 dehydratase 1-like
Gma.3832.2.S1_a_at	0.048852	-36.5603	BU760714	probable mitochondrial 2-oxoglutarate/malate carrier protein-like /// probable mitochondrial 2-oxoglutarate/malate carrier protein-like
Gma.5787.2.S1_x_at	0.006318	-34.8453	CA784580	GDP-mannose 4,6 dehydratase 1-like
GmaAffx.90411.1.S1_x_at	0.008628	-33.5563	CF806136	GDP-mannose 4,6 dehydratase 1-like
Gma.3832.2.S1_x_at	0.030727	-32.4799	BU760714	probable mitochondrial 2-oxoglutarate/malate carrier protein-like
GmaAffx.85640.1.S1_at	0.007514	-31.9695	AW703774	putative nuclease HARBI1-like
Gma.5787.2.S1_s_at	0.009855	-28.6141	CA784580	GDP-mannose 4,6 dehydratase 1-like
GmaAffx.72348.1.S1_at	0.028288	-27.7998	BF324132	---
GmaAffx.87027.1.S1_at	0.006682	-27.1994	AW831345	glucan endo-1,3-beta-glucosidase 12-like
GmaAffx.19194.1.S1_at	7.42E-06	-26.7896	BM568005	transcription factor ORG2-like
Gma.6489.1.S1_at	0.010548	-26.5865	BQ612490	---
Gma.3668.3.S1_at	0.000964	-26.5308	AW704362	1-aminocyclopropane-1-carboxylate oxidase 5-like
Gma.11156.2.S1_s_at	0.002446	-25.3402	BU964892	uncharacterized LOC100782994 /// uncharacterized LOC100810231
GmaAffx.70622.1.S1_at	0.001428	-25.2309	BE806216	auxin-induced protein 5NG4-like
soybean_rRNA_1531_at	0.010652	-24.6732	soybean_rRNA_1531	---
GmaAffx.74021.1.S1_at	0.032406	-24.094	BI972054	proline dehydrogenase 2, mitochondrial-like
GmaAffx.65414.1.S1_at	0.005609	-23.7802	BM094521	---
Gma.10897.1.S1_s_at	0.000914	-23.345	BM519569	uncharacterized LOC100779694 /// uncharacterized LOC100805780
GmaAffx.87722.1.S1_at	2.41E-06	-23.2712	BQ611037	transcription factor ORG2-like

Using similar tools, at 8dpi we were able to identify a total of 3,092 transcripts as being differentially expressed. 2,074 of those transcripts were identified to be up regulated at the same time 1,018 transcripts were known to be down regulated. The transcript of the gene encoding salicylic acid methyl transferase protein (CD416575) increased the most at ~75-fold (see table 5). The most down regulated gene was (BF324269) similar to TR: Q9SMP6 hypothetical 32.2 KD protein with 51-fold lower than in the control (see table 6).

Table 5: List of the 25 most induced genes in soybean (cv.Williams 82) at 8 dpi by *H.glycines* (P-value≤0.05) and (Fold Change>1.5)

Transcript ID	***P*-Value**	**Fold Change**	**Public ID**	**Gene Title**
Gma.12911.1.A1_s_at	0.022089	74.58891	CD416575	Salicylic acid methyl transferase-like protein
Gma.17993.1.S1_at	0.000679	36.99047	BQ628525	protein P21-like
Gma.8768.1.A1_at	0.010103	34.03305	BU764586	MATE efflux family protein FRD3-like
GmaAffx.12883.1.S1_at	0.002374	32.70708	BE820675	WRKY30 protein
GmaAffx.89309.1.S1_at	0.011033	27.30121	CK605560	ruBisCO-associated protein-like
GmaAffx.3364.1.A1_at	0.00113	26.00993	BE347229	serine carboxypeptidase II-3-like
GmaAffx.70875.1.S1_s_at	0.029159	23.9708	AW100693	flavanone 3-dioxygenase-like
Gma.2536.1.S1_at	0.001517	22.45869	BU081272	uncharacterized LOC100500542
Gma.4438.2.S1_a_at	0.043377	22.25167	BQ630502	flavanone 3-dioxygenase-like
Gma.4628.1.S1_at	0.022202	19.73899	BG156501	7-ethoxycoumarin O-deethylase-like
GmaAffx.93343.1.S1_at	0.010223	19.61428	CF809068	probable WRKY transcription factor 70-like
GmaAffx.84058.1.A1_s_at	0.017771	16.4326	CD396207	flavanone 3-dioxygenase-like
Gma.4438.1.S1_at	0.009766	14.54037	CD396138	flavanone 3-dioxygenase-like
Gma.3473.1.S1_at	0.007331	14.05731	BQ628660	17.5 kDa class I heat shock protein-like
GmaAffx.22516.2.S1_at	0.022027	13.47564	BE347576	uncharacterized LOC100797485
GmaAffx.2799.1.A1_at	0.003748	13.13145	BE659055	uncharacterized LOC100527902
GmaAffx.93343.1.S1_s_at	0.001249	12.42393	CF809068	probable WRKY transcription factor 70-like
GmaAffx.41719.1.S1_at	0.001089	11.55013	BM188198	WRKY30 protein
Gma.4438.3.S1_x_at	0.040411	11.54491	BU764535	flavanone 3-dioxygenase-like
Gma.15494.1.S1_at	0.011812	11.41779	BE329760	uncharacterized LOC100306254

GmaAffx.88845.1.S1_at	0.00084	11.06987	CK605838	basic form of pathogenesis-related protein 1-like
GmaAffx.42925.1.S1_s_at	0.019317	10.96173	CA802660	7-ethoxycoumarin O-deethylase-like
Gma.4438.4.S1_x_at	0.038699	10.95126	BE021855	flavanone 3-dioxygenase-like
Gma.9963.1.A1_at	0.004734	10.84107	BE820448	---
GmaAffx.90059.1.S1_s_at	0.000229	10.64669	CF805784	probable leucine-rich repeat receptor-like protein kinase At1g35710-like

Table 6: List of the 25 most suppressed genes in soybean (cv.Williams 82) at 8 dpi by *H.glycines* (P-value≤0.05) and (Fold Change<1.5)

Transcript ID	***P*-Value**	**Fold Change**	**Public ID**	**Gene Title**
Gma.9947.1.S1_at	0.003707	-51.1308	BF324269	uncharacterized LOC100812834
GmaAffx.68621.1.A1_at	0.012642	-24.4784	BI893889	low-temperature-induced 65 kDa protein-like
GmaAffx.92038.1.S1_s_at	0.041456	-19.1112	CF807763	tyrosine aminotransferase
Gma.8530.1.S1_at	0.041402	-18.3312	AF109720.1	seed maturation protein PM35
GmaAffx.37381.1.S1_at	0.019803	-16.0806	BG790575	MYB transcription factor MYB77
Gma.15.1.S1_at	0.039322	-14.3816	U66316.1	desiccation protective protein LEA5
GmaAffx.27550.2.S1_at	0.015102	-13.1066	BQ630398	chloride channel protein CLC-b-like
Gma.8130.1.S1_at	0.030018	-13.0909	AI966352	malate synthase
GmaAffx.10204.1.S1_at	0.021281	-12.9908	BQ298213	---
GmaAffx.75697.1.S1_s_at	0.013206	-12.7112	BU090518	uncharacterized LOC100812834
Gma.10026.1.S1_at	0.011434	-11.7797	BU546301	uncharacterized LOC100306719
GmaAffx.90496.1.S1_s_at	0.017276	-11.2601	CF806221	NAC domain-containing protein 72-like /// NAC domain protein NAC3
GmaAffx.93483.1.S1_s_at	0.041148	-10.4709	CF809208	Xyloglucan endotransglycosylase precursor
GmaAffx.93435.1.S1_at	0.022642	-10.3261	CF809160	uncharacterized LOC100779411
GmaAffx.90968.1.S1_at	0.025894	-10.2509	CF806693	uncharacterized LOC100814726
GmaAffx.11276.1.S1_s_at	0.010925	-9.44138	AW567677	uncharacterized LOC100779411
GmaAffx.86613.2.S1_at	0.039028	-9.37013	BE024005	glutaredoxin-C11-like
Gma.10760.1.S1_at	0.026621	-9.30298	CA937233	ABC transporter G family member 22-like
GmaAffx.92480.1.S1_at	0.034063	-8.82193	CF806037	uncharacterized LOC100811430
Gma.3405.2.S1_s_at	0.024443	-8.5579	BG882930	branched-chain-amino-acid aminotransferase 2, chloroplastic-like
GmaAffx.51821.1.S1_x_at	0.041165	-8.51834	BU761764	Transcriptional factor NAC51
GmaAffx.57970.2.S1_at	0.049848	-8.27232	BF325304	NAC domain protein NAC3
Gma.168.1.S1_at	0.030072	-8.03713	AF117724.1	seed maturation protein PM28
GmaAffx.85721.1.S1_at	0.024506	-7.97094	BM094795	uncharacterized LOC100810548
Gma.4071.1.S1_at	0.031018	-7.9072	BG651692	snakin-2-like

Direct comparisons were made between the two samples (SCN at 8dpi) and (RKN at 12dpi). The analyses focused on the two types of transcripts that were determined to be differentially expressed. We were able to identify a total of 135 transcripts as being present between the 8dpi and 12 dpi and to be up regulated with greater than 1.5-fold change in expression (see table 7)

Table 7: List of the common induced genes between SCN 8dpi and RKN 12dpi

Transcript ID	**Public ID**	**Gene Title**	**12dpi *P*-value**	**12dpi_FC**	**8dpi *P*-value**	**8dpi_FC**
Gma.10033.1	AI855908	---	0.0376028	1.568306	0.0366192	1.636095
Gma.10150.1	AW277321	protein SRG1-like	0.01949003	5.728352	0.03454021	4.312641
Gma.10406.1	BU761235	uncharacterized LOC100775399	0.01257177	2.058265	0.04483251	3.119783
Gma.10523.1	BG550873	uncharacterized LOC100783984	0.000245256	7.840076	0.03547237	3.528934
Gma.10580.1	BF324224	4-hydroxyphenylpyruvate dioxygenase	0.005313644	1.513906	0.0339468	3.594257
Gma.12094.1	CD395088	Uncharacterized LOC100306688	0.00034436	7.180955	0.02843143	2.614575
Gma.12463.1	BE807144	uncharacterized LOC100790222	0.004093942	9.314862	0.04913537	2.521092
Gma.12569.1	CD391541	Uncharacterized LOC100819476	0.003223673	2.076129	0.03590029	2.141525
Gma.13013.2	AW457857	uncharacterized LOC100784533	0.000780007	17.88636	0.03277417	2.899262
Gma.13045.1	CD401609	Purple acid phosphatase 17-like	0.0262264	2.257731	0.03488898	1.900352
Gma.13124.1	CD393402	uncharacterized LOC100819981	0.004458848	2.893063	0.04048982	1.836941
Gma.1339.1	BQ272933	homogentisate phytyltransferase 1, chloroplastic-like	0.000469686	3.036177	0.03896257	2.178197
Gma.13433.1	CD398049	---	0.02125882	4.941759	0.03580524	2.468501
Gma.13476.1	CD398801	superkiller viralicidic activity 2-like 2-like	0.003501847	2.092104	0.04286844	1.836898
Gma.13553.1	CD400444	nudix hydrolase 16, mitochondrial-like	0.00576102	1.93206	0.02430014	1.623014
Gma.1367.1	BQ297577	uncharacterized LOC100809155	0.01906279	4.259956	0.04226048	1.663374
Gma.13766.1	CD405069	transcription factor bHLH63-like	7.97969E-05	2.007474	0.01101826	1.647539
Gma.14016.1	CD413373	Pentatricopeptide repeat-containing protein At5g67570, chloroplastic-like	0.04856291	2.265211	0.03950592	1.532062
Gma.14288.1	CA852004	cationic peroxidase 1-like	0.01161702	5.445443	0.03576035	9.449177
Gma.14515.3	AW458157	Bystin-like	0.008054614	3.614005	0.03975817	1.947921
Gma.14907.1	CD400055	uncharacterized LOC100798718	0.001288054	3.755445	0.04738405	2.036797

Gma.15487.1	D63781.1	epoxide hydrolase	0.04538458	1.530013	0.01950934	2.169333
Gma.15695.1	BE022791	6-phosphofructokinase 2-like	0.003843545	2.60692	0.03148702	2.421559
Gma.16528.1	CD398706	Transcription factor bHLH77-like	0.02869755	1.946698	0.0406879	3.126229
Gma.16573.1	CD402611	uncharacterized LOC100784194	0.02397743	2.500149	0.03488163	2.634429
Gma.16683.1	BG510367	30S ribosomal protein 1, chloroplastic-like	0.003216296	4.063587	0.02409317	2.476659
Gma.16683.1	BG510367	30S ribosomal protein 1, chloroplastic-like	0.000178928	6.090424	0.04280882	2.800231
Gma.16762.1	BU546275	Beta-glucosidase-like SFR2, chloroplastic-like	0.000688517	2.176447	0.02934298	1.528768
Gma.17086.1	AI856208	probable 2-oxoglutarate/Fe(II)-dependent dioxygenase-like	0.01763233	1.839912	0.04162265	5.437317
Gma.17816.1	BI970976	cationic peroxidase 1-like	0.006365458	2.023132	0.03866445	2.231271
Gma.2183.1	CD402949	L-allo-threonine aldolase-like	0.02095573	1.551566	0.02831243	1.611296
Gma.2292.2	BQ297915	uncharacterized LOC100527798	0.000335576	8.721923	0.03535151	7.848803
Gma.2772.1	BG238134	---	0.005085129	1.890942	0.04414454	2.902571
Gma.3394.1	BQ627958	---	0.006981709	3.550197	0.04646153	1.674104
Gma.3409.1	CA936258	non-specific lipid-transfer protein-like	0.01839704	5.168887	0.04552034	2.761917
Gma.3530.1	BU545704	---	0.04105533	1.618805	0.04168939	1.822494
Gma.3766.1	BE023976	uncharacterized LOC100799979	0.002165881	5.484429	0.03330987	2.476669
Gma.3828.1	BE657673	uncharacterized LOC100305963	0.02682123	2.555454	0.03094236	2.548332
Gma.4071.1	BG651692	snakin-2-like	7.21135E-05	4.413213	0.03101785	7.907204
Gma.4100.1	BU549021	basic 7S globulin-like	1.82768E-05	13.95827	0.02956951	3.019708
Gma.4120.1	AW310714	uncharacterized LOC100788174	4.3857E-05	6.471719	0.03726543	2.361957
Gma.4649.1	BQ628783	nicotinamide mononucleotide adenylyltransferase 1-like	0.01319972	1.939536	0.005571368	1.518161
Gma.4868.1	BI969469	Cysteine-rich receptor-like protein kinase 29-like	0.000698906	4.709741	0.01990893	4.33853
Gma.4971.1	BI967269	uncharacterized LOC100776551	0.00178417	2.22608	0.04850975	1.500986
Gma.5287.1	CD403895	uncharacterized LOC100784244	0.01198887	1.574375	0.02617877	1.621445
Gma.5582.1	AW598007	neurochondrin-like	0.008542069	1.505062	0.02847868	1.550345
Gma.5622.1	BG363493	uncharacterized LOC100306414	0.01235342	3.467613	0.02972919	5.290913
Gma.5778.1	BM091675	tRNA dimethyl allyl transferase 2-like	0.006375758	1.94444	0.04465096	1.550565
Gma.5928.1	BQ610440	translation initiation factor eIF-2B subunit delta-like	0.0207281	1.884573	0.04569256	1.661134
Gma.6246.1	AW348532	uncharacterized LOC100780772	0.02885914	4.7876	0.02381455	2.951122
Gma.6707.1	CD393246	uncharacterized LOC100806250	0.002523744	1.946266	0.03747033	2.035147

Gma.6893.1	CD418129	uncharacterized LOC100800367	0.000471427	1.824626	0.03758242	1.51488
Gma.6942.1	BE802944	Probable tyrosine-protein phosphatase At1g05000-like	0.000863207	3.594007	0.02894507	2.628062
Gma.7220.1	AI495461	D-aminoacyl-tRNA deacylase-like	0.00085937	5.555176	0.03259616	1.754763
Gma.7241.1	CD418665	uncharacterized LOC100500622	0.000309329	10.35495	0.04716657	3.579052
Gma.7267.2	BM886781	zinc finger protein ZPR1-like	0.003288627	1.586612	0.0407183	1.620353
Gma.7326.1	AW277720	probable adenylate kinase isoenzyme 6-like	0.01130452	2.888915	0.04778237	1.762135
Gma.8023.1	AW704826	uncharacterized LOC100799644	0.01214032	2.724444	0.02734742	1.677046
Gma.8257.1	AW350412	uncharacterized LOC100789682	0.02311344	3.700484	0.02892661	1.570319
Gma.8669.1	BI967417	1,4-alpha-glucan-branching enzyme-like	0.003151311	3.231646	0.03249377	2.36812
Gma.8760.1	BU764000	jmjC domain-containing protein 7-like	0.01324339	2.379353	0.04562952	1.819638
Gma.9572.1	BU550343	poly(A)-specific ribonuclease PARN-like	0.01009862	3.944811	0.0316271	1.752537
Gma.961.1	BE824056	DNA (cytosine-5)-methyltransferase DRM1-like	0.001188416	3.235985	0.02073403	1.673817
Gma.9726.1	CA936091	cysteine-rich receptor-like protein kinase	0.000154723	5.844724	0.03383869	1.557216
GmaAffx.1027.1	BU548422	U3 small nucleolar RNA-associated protein 15 homolog	0.02607161	2.183562	0.0187196	2.019908
GmaAffx.1037.1	BU549658	ATP-dependent zinc metalloprotease FtsH 1-like	0.002481519	2.946541	0.01541169	2.284206
GmaAffx.1065.1	AW350646	TVP38/TMEM64 family membrane protein slr0305-like	0.000794814	1.993828	0.0375825	1.871492
GmaAffx.1097.1	BE657914	---	0.01171016	5.399883	0.03512667	2.808498
GmaAffx.13133.1	BQ628177	---	0.01224006	2.684912	0.03119445	2.332567
GmaAffx.15779.1	CD403404	uncharacterized LOC100801047	0.002203217	2.315775	0.0186687	2.298334
GmaAffx.22155.1	BE823381	zinc finger protein VAR3, chloroplastic-like	0.001508433	3.237245	0.03888601	1.953012
GmaAffx.25641.1	BF598552	ethylene-responsive transcription factor ERF113-like	4.02166E-05	6.769797	0.006630147	1.564228
GmaAffx.25840.1	CD402384	Photosystem I reaction center subunit VI, chloroplastic-like	0.001412276	4.273167	0.03149455	2.136441
GmaAffx.26456.1	BE608750	NHL repeat-containing protein 2-like	0.03110995	2.629003	0.02077112	2.563723
GmaAffx.26700.1	BI094913	protein pleiotropic regulatory locus 1-like	0.03098788	2.009905	0.03835554	1.550941
GmaAffx.26739.1	BI315875	Uncharacterized LOC100792596	0.02978001	2.211263	0.04331096	1.654571
GmaAffx.27488.1	AW311080	serine/threonine-protein kinase CTR1-like	0.02680389	1.634126	0.02377344	1.955157
GmaAffx.28511.1	BM953949	uncharacterized LOC100499774	0.003964692	1.909722	0.02126441	1.572589
GmaAffx.29225.3	BG041895	Transcription factor bHLH121-like	0.004890555	1.899933	0.004668455	1.856678
GmaAffx.31232.1	BE823884	CCAAT/enhancer-binding protein zeta-like	0.02172951	2.669481	0.03479784	1.883169
GmaAffx.3327.1	CD400412	small glutamine-rich tetratricopeptide repeat-containing protein alpha-like	0.01831586	5.523249	0.04985942	2.05206

GmaAffx.33428.1	BU548001	meiotic recombination protein SPO11-1-like	0.02457454	1.859865	0.04737999	1.587116
GmaAffx.35506.1	BE820574	uncharacterized LOC100306338 /// uncharacterized LOC100775733	0.00104315	9.106568	0.0412941	3.947585
GmaAffx.36470.1	BU550552	uncharacterized LOC100818144	0.02645044	1.548386	0.041728	1.892275
GmaAffx.39130.1	AW308946	bidirectional sugar transporter SWEET14-like	0.02323845	2.306602	0.03781219	3.983806
GmaAffx.39591.1	AW759446	Uncharacterized LOC100810685	0.03654424	1.829951	0.02635753	1.599699
GmaAffx.40889.1	BG507313	---	0.006833449	2.958084	0.04617156	2.251881
GmaAffx.41218.1	BI470625	---	0.04536197	2.445846	0.04340711	1.598939
GmaAffx.41274.1	BI699087	uncharacterized LOC100781578	0.003886853	3.504569	0.02003873	1.772202
GmaAffx.42886.1	BI702156	rae1-like protein At1g80670-like	0.0111471	2.111374	0.04318198	1.720042
GmaAffx.46379.1	BM521726	outer envelope protein of 80 kDa, chloroplastic-like	0.04508962	1.781122	0.02732947	1.529655
GmaAffx.47232.1	CA935083	oxygen-evolving enhancer protein 1, chloroplastic-like	0.01139103	4.814871	0.04699392	1.90406
GmaAffx.48349.1	AW306256	zinc finger CCCH domain-containing protein 20-like	0.000475752	3.576424	0.03853845	1.611768
GmaAffx.49788.1	BG154410	methyltransferase-like protein 2-like	0.01597592	1.714884	0.005043459	1.549462
GmaAffx.50835.1	AW349835	RING-H2 finger protein ATL3-like	0.001755986	14.86833	0.03830458	2.499539
GmaAffx.53489.1	BE823335	DNA polymerase V-like	0.02510608	2.385188	0.03193736	1.552365
GmaAffx.54553.1	BI469916	uncharacterized LOC100819128	0.002653311	7.998586	0.03030778	2.214723
GmaAffx.55452.1	BM309357	glutathione S-transferase GST 24	0.04522489	2.009984	0.04345685	3.248652
GmaAffx.56800.1	CD394842	polyneuridine-aldehyde esterase-like	0.000231216	6.033762	0.02519215	1.857365
GmaAffx.60673.1	BU080770	F-box/kelch-repeat protein At3g23880-like	0.04744348	4.048864	0.02212726	6.261647
GmaAffx.60786.1	BU547092	uncharacterized LOC100778651	0.02739487	2.574627	0.04903957	1.638829
GmaAffx.61126.1	BI321919	uncharacterized LOC100785868	0.03681396	2.65422	0.04951972	4.348725
GmaAffx.6182.1	BE610496	---	0.02421321	2.230978	0.04073717	2.676402
GmaAffx.64000.1	BU762209	uncharacterized LOC100806374	0.02288943	4.46226	0.0415932	2.606523
GmaAffx.64236.1	AI960326	purple acid phosphatase 3-like	0.000440445	7.192724	0.04386033	5.289434
GmaAffx.65609.1	BG652296	two-component response regulator-like APRR1-like	1.65286E-06	3.4888	0.04478869	1.679935
GmaAffx.68386.1	BU549170	uncharacterized LOC100305943	0.004075633	1.516721	0.02026853	3.526421
GmaAffx.69171.1	BM525132	---	0.02880798	3.069581	0.04843928	1.611701
GmaAffx.69354.1	BM892381	---	0.004321404	3.078672	0.02160577	2.551687
GmaAffx.69955.1	BU578900	protein FAR1-RELATED SEQUENCE 5-like	0.009248596	2.54344	0.02333038	1.564834
GmaAffx.71293.1	AW569180	uncharacterized LOC100780772	0.001194401	5.018656	0.004811007	1.921461

GmaAffx.72140.1	BU549812	DEAD-box ATP-dependent RNA helicase ISE2, chloroplastic-like	0.00155459	1.889539	0.03615095	1.555026
GmaAffx.74261.1	BQ741802	RING-H2 finger protein ATL78-like	0.006124046	1.849354	0.02169378	2.877942
GmaAffx.74525.2	BI892728	WRKY6	0.000672384	4.967464	0.03945706	4.187678
GmaAffx.75675.1	AW351172	phenylalanine ammonia-lyase class 3-like	0.000530387	6.949903	0.04706874	1.924545
GmaAffx.76678.1	BE823183	putative glucose-6-phosphate 1-epimerase-like	0.03977728	3.195505	0.03979145	2.271203
GmaAffx.783.1	BE822842	uncharacterized LOC100798195	0.008879902	1.882612	0.02824946	2.245803
GmaAffx.79552.1	CD408681	Uncharacterized LOC100777466	0.00712335	3.90356	0.03297941	2.014489
GmaAffx.8102.1	AW156767	uncharacterized LOC100817132	0.02768882	1.973472	0.04462853	1.814815
GmaAffx.8111.1	BE659080	tropinone reductase homolog At1g07440-like	0.01964935	1.813208	0.03408654	4.127934
GmaAffx.8247.2	BU550218	eukaryotic initiation factor iso-4F subunit p82-34-like	0.0267294	2.374767	0.003845223	1.510933
GmaAffx.8278.1	BE820944	probable alpha-mannosidase I MNS5-like	0.0372899	2.658956	0.03621329	1.724456
GmaAffx.86081.1	BG041479	uncharacterized LOC100795813	0.002824626	2.099602	0.03631355	1.534298
GmaAffx.86185.1	CA803005	---	0.006626072	1.544576	0.04502255	1.637664
GmaAffx.86613.2	BE024005	glutaredoxin-C11-like	0.01062685	4.359408	0.03902806	9.370126
GmaAffx.87343.1	BM524036	uncharacterized LOC100804440	0.03437921	1.593393	0.03717519	1.841479
GmaAffx.8868.1	BU548596	pentatricopeptide repeat-containing protein At3g04760, chloroplastic-like	0.02809853	2.946599	0.0462669	1.503176
GmaAffx.9069.1	BG881714	uncharacterized LOC100500148	0.01011165	1.808228	0.02674786	1.970166
GmaAffx.90984.1	CF806709	Germin-like protein subfamily 1 member 7-like	0.01394249	3.505045	0.002679921	3.897544
GmaAffx.91358.1	CF807083	uncharacterized LOC100785040	0.02880331	1.877666	0.01984674	5.107628
GmaAffx.91805.1	CF807530	GDSL esterase/lipase 5-like	1.98193E-05	59.54299	0.02557859	4.197063
GmaAffx.92875.1	CF808600	ribonuclease J-like	0.000672941	2.111578	0.03647465	1.625508
GmaAffx.92888.1	CF808613	probable mitochondrial chaperone BCS1-B-like	0.02801011	3.310903	0.02353319	5.537118
HgAffx.20336.3	CA940243	---	0.01133449	9.432377	0.0286032	14.26889
HgAffx.21442.1	BI748311	---	0.03934574	1.636702	0.03354049	1.964414

Similarly, we identified a total of 82 transcripts as being present between the 8dpi and 12 dpi and to be down regulated with less than 1.5-fold change in expression (see table 8)

Table 8: List of the common suppressed genes between SCN 8dpi and RKN 12dpi

ProbeSet ID	Public ID	Gene title	8dpi *P*-Value	8dpi_FC	12dpi *P*-Vlaue	12dpi_FC
Gma.10231.1.A1_at	CF920420	pistil-specific extensin-like protein-like	0.03947296	-1.518521	0.00012623	-8.05398
Gma.10231.1.A1_s_at	CF920420	pistil-specific extensin-like protein-like	0.03987376	-1.687837	0.00131508	-11.1206
Gma.10854.1.S1_at	CA802334	auxin-induced protein 5NG4-like	0.01368124	-5.687873	0.01966562	-1.95295
Gma.10911.1.S1_at	AW396425	Carbonic anhydrase, chloroplastic-like	0.0435969	-2.98565	0.003594878	-2.76763
Gma.12121.1.S1_at	BU550753	uncharacterized LOC100526906	0.03452417	-5.350936	0.03600441	-2.68421
Gma.12761.2.S1_a_at	BI974834	beta-carotene 3-hydroxylase, chloroplastic-like	0.04513256	-2.384471	0.001165967	-1.61117
Gma.13166.1.S1_at	BQ296064	probable cytokinin riboside 5'-monophosphate phosphoribohydrolase LOGL6-like	0.04750408	-1.629183	0.007276475	-3.92655
Gma.13227.1.A1_at	BE610302	LOB domain-containing protein 39-like	0.03256194	-1.752009	0.002865424	-1.72359
Gma.13307.1.S1_at	BQ628526	uncharacterized LOC100816033	0.03019308	-1.811931	0.03908329	-5.98826
Gma.13394.1.A1_x_at	CD391087	uncharacterized LOC100798343	0.03103626	-1.980799	0.003308422	-1.67056
Gma.15.1.S1_at	U66316.1	desiccation protective protein LEA5	0.03932206	-14.38161	0.04272985	-4.38748
Gma.15247.2.A1_at	AW706289	Uncharacterized LOC100804534	0.02089023	-1.527274	0.00863161	-5.37063
Gma.15725.1.S1_at	BQ627887	uncharacterized LOC100526935	0.03977346	-2.167654	0.03285109	-9.23956
Gma.1583.1.S1_at	AW568736	uncharacterized LOC100807834	0.03593807	-1.681375	0.000179979	-6.97036
Gma.1583.1.S1_s_at	AW568736	uncharacterized LOC100786628 /// uncharacterized LOC100807834	0.002367533	-1.77601	0.006295417	-6.66879
Gma.1619.1.S1_at	AW472195	uncharacterized LOC100775453	0.01838425	-1.674145	0.03712733	-2.16016
Gma.1871.2.S1_a_at	BF219517	protein notum homolog /// protein notum homolog	0.03639354	-2.499103	0.0352451	-1.79703
Gma.1906.1.S1_at	BI095213	polyol transporter 5-like	0.03286872	-2.521651	0.004433471	-3.9141
Gma.1991.1.S1_at	BM521488	magnesium transporter NIPA2-like	0.04405863	-1.801732	0.02687277	-2.77792
Gma.2044.2.S1_at	BE821230	abscisic stress ripening-like protein	0.03971772	-3.190765	0.01694225	-7.34524
Gma.2195.1.S1_at	BE823097	beta-galactosidase 1-like	0.01530426	-2.480821	0.000358533	-1.58812
Gma.2560.3.S1_a_at	AI938055	uncharacterized LOC100783794 /// uncharacterized LOC100791877	0.02128738	-1.527765	0.01439303	-1.71475
Gma.2777.1.S1_at	BQ785406	trihelix transcription factor GT-2-like	0.04589326	-1.686584	0.03191457	-1.84973
Gma.3232.1.S1_at	CD405154	receptor-like protein kinase	0.04642637	-1.720302	0.04679602	-1.96097
Gma.3250.1.S1_a_at	BU545972	fructose-1,6-bisphosphatase, cytosolic-like	0.02108962	-2.676015	0.04511904	-2.13072
Gma.4372.1.S1_at	AI443505	momilactone A synthase-like	0.03597546	-1.965622	0.003410752	-2.88354
Gma.4719.1.A1_at	AW309958	Uncharacterized LOC100818256	0.04500378	-2.401937	0.003504482	-16.8566

Gma.5157.2.A1_s_at	BI971140	Uncharacterized LOC100527823	0.001399111	-3.53494	0.03911397	-2.62723
Gma.5799.2.S1_at	BM523744	uncharacterized LOC100797118	0.034429	-2.176139	0.00919764	-3.2945
Gma.620.1.S1_at	CA820071	glutathione transferase	0.03562476	-2.5975	0.04196642	-2.26642
Gma.7580.1.S1_at	BE057763	uncharacterized LOC100306576	0.004191696	-5.330752	0.000440682	-2.0653
Gma.8072.2.A1_at	L01449.1	G-box binding factor	0.02447939	-1.541112	0.008360478	-2.06874
Gma.8372.2.S1_at	BM890886	dehydration-responsive element binding protein 2	0.008316709	-2.599642	0.01677373	-6.52948
Gma.8530.1.S1_at	AF109720.1	seed maturation protein PM35	0.04140187	-18.33123	0.00786423	-3.08583
Gma.9559.1.S1_x_at	BU548270	uncharacterized LOC100779660	0.006989331	-1.500811	0.03212912	-1.64731
GmaAffx.10497.1.S1_at	AW201969	alpha-aminoadipic semialdehyde synthase-like	0.02472744	-3.928156	0.000199779	-1.98285
GmaAffx.13228.1.S1_at	BG043416	protein TRANSPARENT TESTA 12-like	0.03574866	-1.895989	0.007646346	-5.09852
GmaAffx.15390.1.S1_at	BU927171	uncharacterized LOC100527684	0.0307361	-2.872873	0.000883194	-3.87725
GmaAffx.15904.1.S1_at	AW348412	UPF0098 protein MTH_273-like	0.0111549	-5.132559	0.002682984	-2.29735
GmaAffx.16271.1.S1_at	CA936304	peroxidase 21-like	0.02975406	-1.818123	8.16E-05	-1.9184
GmaAffx.17681.1.S1_at	BF425844	aspartic proteinase nepenthesin-1-like	4.78E-02	-2.501568	0.0105158	-4.08376
GmaAffx.19398.1.S1_at	AW278904	putative oxidoreductase C1F5.03c-like	0.03895414	-1.74483	0.006469268	-1.86002
GmaAffx.20978.1.S1_at	BE058691	receptor-like protein kinase	0.04265531	-1.657352	0.02352995	-1.5597
GmaAffx.21184.1.S1_at	BU547236	uncharacterized LOC100784515	0.02545501	-1.807712	0.03094657	-4.4195
GmaAffx.23831.1.S1_at	BM731488	---	0.04076584	-1.511095	0.0091895	-2.27465
GmaAffx.28080.1.S1_at	BM094868	xyloglucan galactosyltransferase KATAMARI1 homolog	0.01871661	-1.70242	0.008481944	-1.78892
GmaAffx.29093.1.S1_at	BU761457	common plant regulatory factor 1-like	0.03096743	-3.483354	0.003548778	-2.06179
GmaAffx.32206.1.S1_at	BF595187	uncharacterized LOC100811193	1.80E-05	-1.911724	2.94E-05	-2.69085
GmaAffx.32345.1.A1_s_at	BI970277	putative pectinesterase/pectinesterase inhibitor 24-like	1.85E-02	-1.92E+00	0.000381555	-6.21548
GmaAffx.36084.2.S1_at	BM187652	auxin-induced protein 5NG4-like	0.04809057	-3.517916	0.022078	-1.94633
GmaAffx.41479.1.A1_at	BI969388	Uncharacterized LOC100527246	0.004973411	-1.510831	0.03441413	-1.51592
GmaAffx.45506.1.S1_at	AI736732	uncharacterized LOC100786288	0.0380246	-1.818308	0.04375289	-2.44636
GmaAffx.46427.1.S1_at	BM525486	probable serine/threonine-protein kinase At5g41260-like	0.0138231	-2.379946	0.006296345	-1.96894
GmaAffx.55182.1.S1_at	AW704358	uncharacterized LOC100811678	0.01865639	-1.576757	0.01774895	-2.93547
GmaAffx.60481.1.S1_at	BE210582	uncharacterized LOC100803116	0.02583212	-1.543211	0.006144269	-2.3205
GmaAffx.62936.1.S1_at	BQ094863	Ninja-family protein AFP3-like	0.04264766	-2.292746	0.04435145	-4.43405
GmaAffx.63207.1.S1_at	BF070134	uncharacterized LOC100802385	0.02921411	-1.705447	0.000766625	-77.1037
GmaAffx.67401.1.S1_at	BM528173	uncharacterized LOC100792129 /// glycogenin-1-like	0.02906725	-5.812291	0.002449928	-2.90303
GmaAffx.67834.1.S1_at	BM188411	GDSL esterase/lipase At5g55050-like	0.03370081	-1.515373	0.03031962	-7.24755
GmaAffx.71646.1.S1_at	BE057187	protein PLANT CADMIUM RESISTANCE 8-like	0.02010136	-2.796984	0.03351277	-1.70111
GmaAffx.71698.1.S1_at	BQ297799	Uncharacterized LOC100815067	0.0278759	-1.526376	0.02976543	-1.82126
GmaAffx.72348.1.S1_at	BF324132	---	0.0362578	-4.908624	0.02828812	-27.7998
GmaAffx.78720.1.S1_at	BM526650	protein notum homolog	0.007026627	-3.254883	0.000287395	-4.1669
GmaAffx.80465.1.S1_at	BU548370	glycogenin-1-like	0.03694599	-4.563055	0.000654154	-1.8883
GmaAffx.81045.1.A1_at	BE659926	Nuclear transcription factor Y subunit C-9-like	0.02878319	-1.530354	0.0160713	-1.85907

GmaAffx.81851.1.S1_at	AW706751	uncharacterized LOC100306030	0.01452069	-2.373359	0.000618206	-2.33072
GmaAffx.84528.1.S1_at	BG882931	uncharacterized LOC100816033	0.01976846	-1.893754	0.000845698	-16.9286
GmaAffx.86997.1.A1_at	BI970908	endoglucanase 5-like	0.009023527	-1.765154	0.001556493	-3.40082
GmaAffx.87092.1.S1_s_at	BQ452697	U-box domain-containing protein 25-like	0.01676495	-1.566993	0.041052	-2.79743
GmaAffx.87939.1.S1_at	BU090479	F-box/FBD/LRR-repeat protein At5g56420-like	0.004484463	-1.792337	0.02555044	-1.96402
GmaAffx.88781.1.S1_s_at	CK605907	uncharacterized LOC100527907	0.04858199	-1.75434	0.008664452	-5.92668
GmaAffx.88862.1.A1_at	CK605821	Uncharacterized LOC100527308	0.01054955	-2.20485	0.002012433	-1.83036
GmaAffx.88955.1.A1_at	CK605728	uncharacterized LOC100527823	6.05E-05	-2.73758	0.000357394	-3.87425
GmaAffx.89396.1.S1_x_at	CK605572	probable xyloglucan endotransglucosylase/hydrolase protein 26-like	0.01190312	-1.72E+00	0.000351033	-6.91432
GmaAffx.90646.1.S1_at	CF806371	BTB/POZ domain-containing protein At2g30600-like	0.01748964	-2.431054	0.005234999	-5.67861
GmaAffx.90929.1.A1_at	CF806654	Uncharacterized LOC100527212	0.02816777	-2.883007	0.00155364	-3.26291
GmaAffx.90942.1.S1_at	CF806667	GTP-binding protein YPTM2-like	0.04963734	-1.7202	0.0451893	-3.14424
GmaAffx.91248.1.S1_s_at	CF806973	protein phosphatase 2C 37-like	0.04334792	-3.099756	0.001133694	-4.63677
GmaAffx.91960.1.S1_s_at	CF807685	2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial-like	0.01054893	-5.846241	0.001466255	-1.80391
GmaAffx.92834.1.S1_at	CF808559	uncharacterized LOC100305493	0.02449094	-3.229797	0.02620623	-2.09973
GmaAffx.92834.1.S1_x_at	CF808559	uncharacterized LOC100305493	0.03054195	-3.120372	0.0357617	-2.06163
GmaAffx.93189.1.S1_at	CF808914	beta-ureidopropionase-like	0.03263654	-2.155522	0.03164625	-2.0139

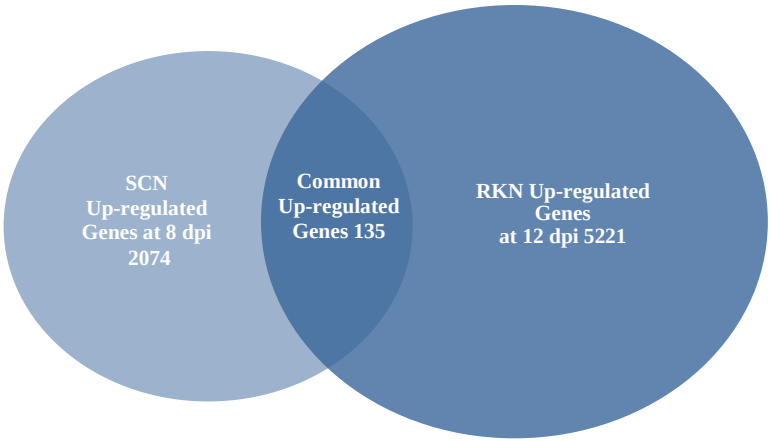


Figure 15: Venn diagram depicting up regulated genes for RKN 12dpi versus SCN 8dpi. The three portions of the Venn diagram show the total number of induced transcripts in RKN at 12dpi; total number of induced transcripts in SCN at 8dpi; the common induced transcripts in both samples.

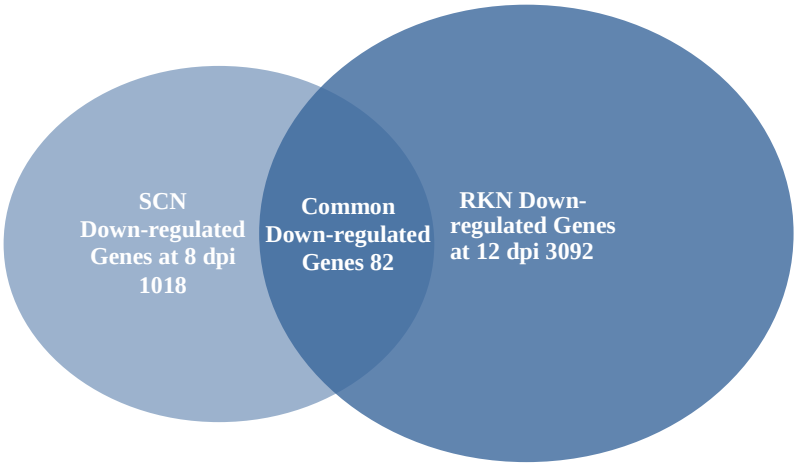


Figure 16: Venn diagram depicting down regulated genes for RKN 12dpi versus SCN 8dpi. The three portions of the Venn diagram show the total number of suppressed transcripts in RKN at 12dpi; total number of suppressed transcripts in SCN at 8dpi; the common suppressed transcripts in both samples.

When the GO annotation became available through B2G, we uploaded the existing annotation file to the application. We performed direct statistical analysis on gene function information. A common analysis is the statistical assessment of GO term enrichments in a group of interesting genes when compared with a reference group. Fisher's Exact Test was performed applying robust FDR (false discovery rate) and *P*-value correction for multiple testing and returns a list of significant GO terms ranked by their *P*-values. In our case, we compared the SCN list of the induced genes as the group of interesting genes to the RKN list of the induced genes as the reference group (see table 9).

Table 9: List of GO terms for the induced genes in SCN that are significantly different from RKN.

'F' stands for molecular function while "P" stands for biological process.

GO-ID	Term	Category	FDR	P-Value
GO:0016706	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	F	0.02149	7.03E-06
GO:0016491	oxidoreductase activity	F	0.028676	1.88E-05
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	F	0.035043	3.75E-05
GO:0005506	iron ion binding	F	0.035043	5.25E-05
GO:0055114	oxidation-reduction process	P	0.035043	5.89E-05
GO:0045486	naringenin 3-dioxygenase activity	F	0.035043	7.57E-05
GO:0009620	response to fungus	P	0.035043	8.02E-05
GO:0003824	catalytic activity	F	0.089722	2.59E-04
GO:0004185	serine-type carboxypeptidase activity	F	0.089722	2.93E-04
GO:0070008	serine-type exopeptidase activity	F	0.089722	2.93E-04
GO:0009813	flavonoid biosynthetic process	P	0.093084	3.35E-04
GO:0009617	response to bacterium	P	0.097735	3.86E-04
GO:0020037	heme binding	F	0.097735	4.16E-04
GO:0009812	flavonoid metabolic process	P	0.117915	5.40E-04
GO:0004180	carboxypeptidase activity	F	0.144894	7.11E-04
GO:0046906	tetrapyrrole binding	F	0.155384	8.41E-04
GO:0006725	cellular aromatic compound metabolic process	P	0.155384	8.79E-04
GO:0019748	secondary metabolic process	P	0.155384	9.15E-04
GO:0051707	response to other organism	P	0.1684	0.001114
GO:0006575	cellular modified amino acid metabolic process	P	0.1684	0.001114
GO:0051704	multi-organism process	P	0.1684	0.001157

GO:0050896	response to stimulus	P	0.201472	0.001479
GO:0009698	phenylpropanoid metabolic process	P	0.201472	0.001516
GO:0060862	negative regulation of floral organ abscission	P	0.221074	0.001808
GO:0060860	regulation of floral organ abscission	P	0.221074	0.001808
GO:0042221	response to chemical stimulus	P	0.23758	0.002021
GO:0042398	cellular modified amino acid biosynthetic process	P	0.265123	0.002342
GO:0009699	phenylpropanoid biosynthetic process	P	0.324793	0.002975
GO:0008152	metabolic process	P	0.32835	0.003115
GO:0006979	response to oxidative stress	P	0.39494	0.004134
GO:0016684	oxidoreductase activity, acting on peroxide as acceptor	F	0.39494	0.004134
GO:0004601	peroxidase activity	F	0.39494	0.004134
GO:0008238	exopeptidase activity	F	0.413271	0.005267
GO:0010942	positive regulation of cell death	P	0.413271	0.005272
GO:0009690	cytokinin metabolic process	P	0.413271	0.005272
GO:0034754	cellular hormone metabolic process	P	0.413271	0.005272
GO:0031349	positive regulation of defense response	P	0.413271	0.005272
GO:0031347	regulation of defense response	P	0.413271	0.005272
GO:0019139	cytokinin dehydrogenase activity	F	0.413271	0.005272
GO:0016209	antioxidant activity	F	0.441116	0.005772
GO:0009607	response to biotic stimulus	P	0.45875	0.006153
GO:0009664	plant-type cell wall organization	P	0.514515	0.007293
GO:0051239	regulation of multicellular organismal process	P	0.514515	0.007406
GO:0004713	protein tyrosine kinase activity	F	0.514515	0.007406
GO:0046914	transition metal ion binding	F	0.644516	0.009518
GO:0010227	floral organ abscission	P	0.644516	0.010251
GO:0005199	structural constituent of cell wall	F	0.644516	0.010251
GO:0009838	Abscission	P	0.644516	0.010251
GO:0019438	aromatic compound biosynthetic process	P	0.644516	0.010331
GO:0050793	regulation of developmental process	P	0.654766	0.010709
GO:0048437	floral organ development	P	0.753093	0.012564
GO:0043169	cation binding	F	0.874857	0.015168
GO:0043167	ion binding	F	0.874857	0.015168
GO:0080134	regulation of response to stress	P	0.890893	0.016611
GO:0010941	regulation of cell death	P	0.890893	0.016611
GO:0016645	oxidoreductase activity, acting on the CH-NH group of donors	F	0.890893	0.016611
GO:0048584	positive regulation of response to stimulus	P	0.890893	0.016611
GO:0006468	protein phosphorylation	P	0.995407	0.018946
GO:0071669	plant-type cell wall organization or biogenesis	P	0.995407	0.019537
GO:0048610	cellular process involved in reproduction	P	0.995407	0.019537
GO:0006950	response to stress	P	1	0.022022
GO:0016709	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NADH or NADPH as one donor, and incorporation of one atom of oxygen	F	1	0.024228
GO:0004497	monooxygenase activity	F	1	0.025796
GO:0050660	flavin adenine dinucleotide binding	F	1	0.028256
GO:0050794	regulation of cellular process	P	1	0.030371
GO:0015893	drug transport	P	1	0.032985
GO:0042493	response to drug	P	1	0.032985
GO:0022603	regulation of anatomical structure morphogenesis	P	1	0.032985
GO:0048638	regulation of developmental growth	P	1	0.032985
GO:0050789	regulation of biological process	P	1	0.034478

GO:0043687	post-translational protein modification	P	1	0.038173
GO:0017171	serine hydrolase activity	F	1	0.038716
GO:0008236	serine-type peptidase activity	F	1	0.038716
GO:0019219	regulation of nucleobase-containing compound metabolic process	P	1	0.041632
GO:0007626	locomotory behavior	P	1	0.042731
GO:0015850	organic alcohol transport	P	1	0.042731
GO:0051865	protein autoubiquitination	P	1	0.042731
GO:0008661	1-deoxy-D-xylulose-5-phosphate synthase activity	F	1	0.042731
GO:0047918	GDP-mannose 3,5-epimerase activity	F	1	0.042731
GO:0045730	respiratory burst	P	1	0.042731
GO:0042447	hormone catabolic process	P	1	0.042731
GO:0002679	respiratory burst involved in defense response	P	1	0.042731
GO:0009631	cold acclimation	P	1	0.042731
GO:0009610	response to symbiotic fungus	P	1	0.042731
GO:0009608	response to symbiont	P	1	0.042731
GO:0016711	flavonoid 3'-monooxygenase activity	F	1	0.042731
GO:0010183	pollen tube guidance	P	1	0.042731
GO:0042330	Taxis	P	1	0.042731
GO:0006935	Chemotaxis	P	1	0.042731
GO:0008493	tetracycline transporter activity	F	1	0.042731
GO:0047134	protein-disulfide reductase activity	F	1	0.042731
GO:0046677	response to antibiotic	P	1	0.042731
GO:0016668	oxidoreductase activity, acting on a sulfur group of donors, NAD or NADP as acceptor	F	1	0.042731
GO:0015520	tetracycline:hydrogen antiporter activity	F	1	0.042731
GO:0040011	Locomotion	P	1	0.042731
GO:0042895	antibiotic transporter activity	F	1	0.042731
GO:0050598	taxane 13-alpha-hydroxylase activity	F	1	0.042731
GO:0042891	antibiotic transport	P	1	0.042731
GO:0004611	phosphoenolpyruvate carboxykinase activity	F	1	0.042731
GO:0000234	phosphoethanolamine N-methyltransferase activity	F	1	0.042731
GO:0008964	phosphoenolpyruvate carboxylase activity	F	1	0.042731
GO:0050918	positive chemotaxis	P	1	0.042731
GO:0015307	drug:hydrogen antiporter activity	F	1	0.042731
GO:0015904	tetracycline transport	P	1	0.042731
GO:0003968	RNA-directed RNA polymerase activity	F	1	0.042731
GO:0009823	cytokinin catabolic process	P	1	0.042731
GO:0010876	lipid localization	P	1	0.042772
GO:0009751	response to salicylic acid stimulus	P	1	0.042772
GO:0016641	oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	F	1	0.042772
GO:0006869	lipid transport	P	1	0.042772
GO:0004683	calmodulin-dependent protein kinase activity	F	1	0.042772
GO:0040008	regulation of growth	P	1	0.042772
GO:0051171	regulation of nitrogen compound metabolic process	P	1	0.043343
GO:0004674	protein serine/threonine kinase activity	F	1	0.043515
GO:0048569	post-embryonic organ development	P	1	0.044586
GO:0044283	small molecule biosynthetic process	P	1	0.046586
GO:0010556	regulation of macromolecule biosynthetic process	P	1	0.046914

Then, we compared the SCN list of the suppressed genes as the group of interesting genes to the RKN list of the suppressed genes as the reference group (see table 10).

Table 10: List of GO terms of the suppressed genes in SCN that are significantly different from RKN.

'F' stands for molecular function; 'P' stands for biological process, while 'C' stands for cellular components.

GO-ID	Term	Category	FDR	P-Value
GO:0005622	Intracellular	C	1.03E-05	3.77E-09
GO:0044424	intracellular part	C	2.44E-05	2.07E-08
GO:0043231	intracellular membrane-bounded organelle	C	2.44E-05	3.56E-08
GO:0043227	membrane-bounded organelle	C	2.44E-05	3.56E-08
GO:0043229	intracellular organelle	C	3.47E-05	7.60E-08
GO:0043226	Organelle	C	3.47E-05	7.60E-08
GO:0044444	cytoplasmic part	C	1.04E-04	2.65E-07
GO:0005737	Cytoplasm	C	1.32E-04	3.85E-07
GO:0009536	Plastid	C	3.28E-04	1.08E-06
GO:0005739	Mitochondrion	C	7.22E-04	2.64E-06
GO:0034660	ncRNA metabolic process	P	0.00174	6.99E-06
GO:0000151	ubiquitin ligase complex	C	0.00282	1.24E-05
GO:0003676	nucleic acid binding	F	0.004184	1.99E-05
GO:0006396	RNA processing	P	0.004232	2.16E-05
GO:0009532	plastid stroma	C	0.011785	6.46E-05
GO:0008026	ATP-dependent helicase activity	F	0.015016	9.87E-05
GO:0031461	cullin-RING ubiquitin ligase complex	C	0.015016	9.87E-05
GO:0070035	purine NTP-dependent helicase activity	F	0.015016	9.87E-05
GO:0009507	Chloroplast	C	0.017962	1.25E-04
GO:0009570	chloroplast stroma	C	0.021244	1.55E-04
GO:0044435	plastid part	C	0.024104	1.85E-04
GO:0044434	chloroplast part	C	0.032829	2.64E-04
GO:0004386	helicase activity	F	0.039505	3.32E-04
GO:0034470	ncRNA processing	P	0.062872	5.51E-04
GO:0080008	CUL4-RING ubiquitin ligase complex	C	0.080927	7.47E-04
GO:0006364	rRNA processing	P	0.080927	8.87E-04
GO:0006310	DNA recombination	P	0.080927	8.87E-04
GO:0016072	rRNA metabolic process	P	0.080927	8.87E-04
GO:0042254	ribosome biogenesis	P	0.080927	8.87E-04
GO:0022613	ribonucleoprotein complex biogenesis	P	0.080927	8.87E-04
GO:0032991	macromolecular complex	C	0.102351	0.001159
GO:0034641	cellular nitrogen compound metabolic process	P	0.146059	0.001743
GO:0006807	nitrogen compound metabolic process	P	0.146059	0.00176

GO:0018130	heterocycle biosynthetic process	P	0.174653	0.002169
GO:0006259	DNA metabolic process	P	0.201509	0.002861
GO:0010319	Stromule	C	0.201509	0.00287
GO:0019395	fatty acid oxidation	P	0.201509	0.00287
GO:0034440	lipid oxidation	P	0.201509	0.00287
GO:0006635	fatty acid beta-oxidation	P	0.201509	0.00287
GO:0046483	heterocycle metabolic process	P	0.242152	0.003538
GO:0006399	tRNA metabolic process	P	0.296858	0.004566
GO:0008483	transaminase activity	F	0.296858	0.004566
GO:0015979	Photosynthesis	P	0.296858	0.004716
GO:0006412	Translation	P	0.296858	0.004771
GO:0048608	reproductive structure development	P	0.312797	0.005141
GO:0009793	embryo development ending in seed dormancy	P	0.345131	0.005798
GO:0044464	cell part	C	0.362236	0.00635
GO:0005623	Cell	C	0.362236	0.00635
GO:0043234	protein complex	C	0.369079	0.006605
GO:0006139	nucleobase-containing compound metabolic process	P	0.39503	0.008134
GO:0090304	nucleic acid metabolic process	P	0.39503	0.008516
GO:0044237	cellular metabolic process	P	0.39503	0.008718
GO:0009579	Thylakoid	C	0.39503	0.008756
GO:0016779	nucleotidyltransferase activity	F	0.39503	0.009121
GO:0006418	tRNA aminoacylation for protein translation	P	0.39503	0.009279
GO:0016876	ligase activity, forming aminoacyl-tRNA and related compounds	F	0.39503	0.009279
GO:0016875	ligase activity, forming carbon-oxygen bonds	F	0.39503	0.009279
GO:0004812	aminoacyl-tRNA ligase activity	F	0.39503	0.009279
GO:0003724	RNA helicase activity	F	0.39503	0.009279
GO:0016905	myosin heavy chain kinase activity	F	0.39503	0.009279
GO:0004084	branched-chain-amino-acid transaminase activity	F	0.39503	0.009279
GO:0043039	tRNA aminoacylation	P	0.39503	0.009279
GO:0004034	aldose 1-epimerase activity	F	0.39503	0.009279
GO:0043038	amino acid activation	P	0.39503	0.009279
GO:0016070	RNA metabolic process	P	0.39503	0.009378
GO:0048316	seed development	P	0.460709	0.012612
GO:0009168	purine ribonucleoside monophosphate biosynthetic process	P	0.460709	0.012788
GO:0009167	purine ribonucleoside monophosphate metabolic process	P	0.460709	0.012788
GO:0009161	ribonucleoside monophosphate metabolic process	P	0.460709	0.012788
GO:0009156	ribonucleoside monophosphate biosynthetic process	P	0.460709	0.012788
GO:0009127	purine nucleoside monophosphate biosynthetic process	P	0.460709	0.012788
GO:0009126	purine nucleoside monophosphate metabolic process	P	0.460709	0.012788

GO:0009124	nucleoside monophosphate biosynthetic process	P	0.460709	0.012788
GO:0009123	nucleoside monophosphate metabolic process	P	0.460709	0.012788
GO:0009081	branched chain family amino acid metabolic process	P	0.460709	0.012788
GO:0030258	lipid modification	P	0.460709	0.012788
GO:0016769	transferase activity, transferring nitrogenous groups	F	0.476692	0.013467
GO:0009409	response to cold	P	0.476692	0.01358
GO:0009266	response to temperature stimulus	P	0.524986	0.015148
GO:0003006	developmental process involved in reproduction	P	0.602908	0.017616
GO:0034357	photosynthetic membrane	C	0.615298	0.018203
GO:0009941	chloroplast envelope	C	0.624894	0.018715
GO:0009791	post-embryonic development	P	0.631784	0.019265
GO:0044446	intracellular organelle part	C	0.631784	0.019613
GO:0044422	organelle part	C	0.631784	0.019613
GO:0009526	plastid envelope	C	0.683486	0.021705
GO:0009790	embryo development	P	0.683486	0.022805
GO:0044436	thylakoid part	C	0.683486	0.022805
GO:0005634	Nucleus	C	0.683486	0.02395
GO:0044260	cellular macromolecule metabolic process	P	0.683486	0.024659
GO:0009152	purine ribonucleotide biosynthetic process	P	0.683486	0.026014
GO:0004518	nuclease activity	F	0.683486	0.026014
GO:0009260	ribonucleotide biosynthetic process	P	0.683486	0.026014
GO:0051188	cofactor biosynthetic process	P	0.683486	0.026335
GO:0010154	fruit development	P	0.683486	0.027276
GO:0009534	chloroplast thylakoid	C	0.683486	0.027569
GO:0031984	organelle subcompartment	C	0.683486	0.027569
GO:0031976	plastid thylakoid	C	0.683486	0.027569
GO:0042598	vesicular fraction	C	0.683486	0.029956
GO:0003863	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring) activity	F	0.683486	0.029956
GO:0010287	Plastoglobule	C	0.683486	0.029956
GO:0005254	chloride channel activity	F	0.683486	0.029956
GO:0005253	anion channel activity	F	0.683486	0.029956
GO:0005247	voltage-gated chloride channel activity	F	0.683486	0.029956
GO:0005834	heterotrimeric G-protein complex	C	0.683486	0.029956
GO:0005792	Microsome	C	0.683486	0.029956
GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	F	0.683486	0.029956
GO:0005759	mitochondrial matrix	C	0.683486	0.029956
GO:0006821	chloride transport	P	0.683486	0.029956
GO:0004725	protein tyrosine phosphatase activity	F	0.683486	0.029956
GO:0019253	reductive pentose-phosphate cycle	P	0.683486	0.029956
GO:0019897	extrinsic to plasma membrane	C	0.683486	0.029956

GO:0008308	voltage-gated anion channel activity	F	0.683486	0.029956
GO:0004527	exonuclease activity	F	0.683486	0.029956
GO:0046474	glycerophospholipid biosynthetic process	P	0.683486	0.029956
GO:0051302	regulation of cell division	P	0.683486	0.029956
GO:0008170	N-methyltransferase activity	F	0.683486	0.029956
GO:0008138	protein tyrosine/serine/threonine phosphatase activity	F	0.683486	0.029956
GO:0019685	photosynthesis, dark reaction	P	0.683486	0.029956
GO:0009062	fatty acid catabolic process	P	0.749212	0.033344
GO:0042651	thylakoid membrane	C	0.749212	0.033383
GO:0010467	gene expression	P	0.764505	0.034344
GO:0034061	DNA polymerase activity	F	0.764725	0.034913
GO:0032259	Methylation	P	0.764725	0.034913
GO:0070279	vitamin B6 binding	F	0.816858	0.038188
GO:0051276	chromosome organization	P	0.816858	0.038188
GO:0030170	pyridoxal phosphate binding	F	0.816858	0.038188
GO:0022414	reproductive process	P	0.837958	0.03948
GO:0009535	chloroplast thylakoid membrane	C	0.846624	0.040507
GO:0055035	plastid thylakoid membrane	C	0.846624	0.040507
GO:0006164	purine nucleotide biosynthetic process	P	0.954183	0.046002
GO:0000003	Reproduction	P	0.975787	0.048862
GO:0006470	protein dephosphorylation	P	0.975787	0.049272
GO:0031969	chloroplast membrane	C	0.975787	0.049272
GO:0016311	Dephosphorylation	P	0.975787	0.049272
GO:0045454	cell redox homeostasis	P	0.975787	0.049272
GO:0003735	structural constituent of ribosome	F	0.975787	0.049272

3.6 Discussion

Microarray experiments typically rely on differential expression analysis methods to identify differences in relative levels of gene expression. However, it is possible that very large differences in gene activity are present when the analysis involves comparing gene expression from different experiments at different stages. Our data reflect similar changes in expression of numerous genes involved in different molecular function in the gall of *G. max* infected by SCN at 8dpi and RKN at 12dpi. For example, one gene (BF324224) encodes 4- hydroxyphenylpyruvate dioxygenase (HPPD) that plants utilize to help produce the cofactors plastoquinone and tocopherol which are essential for the plant to survive (Graham R. Moran, 2005). HPPD was increased in expression 1.5-fold at 12dpi and 3.5-fold at 8dpi. Expression of the transcript (BQ272933) of homogentisate phytyltransferase 1 was induced 3.0-fold at RKN 12dpi and 2.2-fold at SCN 8dpi. (HPT) is a key enzyme to regulate tocopherol synthesis during abiotic stress (Collakova and DellaPenna 2003).

Numerous genes involved in the metabolism process specially oxidoreductase enzymes that commonly known as dehydrogenases or oxidases were differentially expressed. This group of enzymes usually utilizes NADP or NAD+ as cofactors and they are involved in the defense mechanism. Those genes were shown to be induced after the infection with SCN and RKN. For example, the gene (AW277321) was increased 5.7-fold at RKN 12dpi and 4.3-fold at SCN 8dpi. By SCN 8dpi several genes from the same family encoding flavanone 3-dioxygenase were highly up-regulated more than RKN at 12dpi, i.e. BE021855 at 11-fold, (CD396207) at 16.4-

fold, and (CD396138) at 14.5-fold. Another gene involved in metabolism (BI967417) encoding 1,4-alpha-glucan-branching enzyme or glycogen branching enzyme (EC 2.4.1.18) that has a potent role in converting glucose to glycogen and glycogen synthase was induced in our experiments 3.2-fold at 12dpi and 2.3-fold at 8dpi. On the other hand, one of the genes (BQ627887) involved in carbohydrate metabolism was shown to be down-regulated in greater abundance at 12dpi by -9.2 fold than at 8dpi by -2.1 fold. This gene encoding 4-alpha-glucanotransferase DPE2 (EC 2.4.1.25) is essential for the cytosolic metabolism of maltose, an intermediate on the pathway by which starch is converted to sucrose in leaves at night.

Two genes involved in the cell wall were shown to be differentially expressed after infection with *M. incognita* and *H. glycines*. Both genes are encoding one member of the peroxidase enzymes family (cationic peroxidase 1) were shown to be up-regulated (CA852004) at 5.4-fold and 9.4-fold, (BI970976) at 2.0-fold and 2.2-fold for RKN 12dpi and SCN 8dpi, respectively. Studies showed that genes encoding peroxidase are induced by pathogen signals (Lamb and Dixon 1997). The expression of gene (AI443505) encoding momilactone A synthase (EC 1.1.1.295) was shown to be down- regulated in both time points, at 12dpi by -2.9 fold and 8dpi by -2.0 fold. momilactone A involved in the defense mechanism of the plant. Momilactone A is produced in response to attack by a pathogen through the perception of elicitor signal molecules such as chitin oligosaccharide, or after exposure to UV irradiation (Atawong A. et al. 2002). The expression of one of the peroxidase family (CA936304) encoding peroxidase 21 (EC 1.11.1.7) was down-regulated at 12dpi by -1.9 fold and 8dpi by -1.8 fold. Atperox P21 participates in biosynthesis and

degradation of lignin, response to environmental stresses such as wounding, pathogen attack and oxidative stress. These functions might be dependent on each isozyme/isoform in each plant tissue. It might function as heat shock-like defense protein, and may be implicated in the systemic acquired resistance response.

Gene (BU549658) encodes ATP-dependent zinc metalloprotease FtsH 1(AtFTSH1) and [EC 3.4.24.] was up-regulated 2.9-fold at RKN 12dpi and 2.3 at SCN 8dpi. This gene involved in the chloroplast thylakoid formation and in the removal of damaged D1 protein in the photosystem II, preventing cell death under high-intensity light conditions (Lindahl *et al*. 2000). Also, gene (BU549812) encoding DEAD-box ATP-dependent RNA helicase ISE2, chloroplastic (EC 3.6.4.13) was induced by 1.9-fold at 12dpi and 1.5-fold at 8dpi. It is an essential protein required during embryogenesis and involved in post-transcriptional gene silencing. It is necessary for normal plasmodesmata (PD) development and aperture regulation. Gene (CF807530) encodes GDSL esterase/lipase 5-like protein was up-regulated 59.8-fold at RKN 12dpi and 4.2 at SCN 8dpi. The role for the GDSL lipase-like is to catalyze hydrolysis of sinapine during seed germination, leading, via 1-O-sinapoyl-b-glucose, to sinapoyl-L-malate in the seedlings (Claub *et al*. 2008). Another catalytic gene (AW351172) was induced 6.9-fold at 12dpi and 1.9-fold at 8dpi. This gene that encodes phenylalanine ammonia-lyase class 3-like (EC 4.3.1.24) is a member of the ammonia lyase family. PAL involved in the biosynthesis of polyphenol compounds such as flavonoids, phenylpropanoids, and lignin in plants. The activity of PAL is induced dramatically in response to various stimuli such as tissue wounding, pathogenic attack, light, low temperatures, and hormones (Sarkissian

and Gámez 2005). Another member of the lyase family (AW351172) encoding phenylalanine ammonia-lyase class 3-like (EC 4.3.1.24) that is the key enzyme of plant metabolism catalyzing the first reaction in the biosynthesis from L-phenylalanine of a wide variety of natural products based on the phenylpropane skeleton was induced by 6.9-fold at 12dpi and 1.9-fold at 8dpi.

The transcript of the catalytic gene (D63781.1) that encodes epoxide hydrolase was up-regulated at 1.5-fold and 2.1-fold. Epoxide hydrolases (EHs) are enzymes in addition to water have effective biological activities, such as host defenses, control of development, and regulation of inflammation (Newman *et al*. 2005).

Transcript (BM091675) is one of the IPP transferase family that encodes tRNA dimethylallyltransferase 2 (EC 2.5.1.75) was induced 1.9-fold at RKN 12dpi and 1.5-fold at SCN 8dpi. Isopentenyl-diphosphate or tRNA isopentenyltransferase 2 catalyzes the transfer of a dimethylallyl group onto the adenine at position 37 in tRNAs that read codons beginning with uridine, leading to the formation of N6-(dimethylallyl) adenosine (i6A) Involved in the cis-type cytokinin biosynthesis (Miyawaki *et al.* 2006).

The expression of numerous genes encoding cysteine-rich receptor protein kinase (NCRK) that interact with Rop GTPases as central regulators of diverse signaling pathways in plant growth and pathogen defense reflect was altered. For example, gene (BI969469) was up-regulated 4.7-fold and 4.3-fold at 12dpi and 8dpi, respectively. Also, gene (CA936091) was induced 5.8-fold and 1.5-fold.

Our data reflect changes in expression of numerous genes involved in the plant defense system in the RKN gall at 12dpi and SCN gall 8dpi. For example, one gene was increased at 2.0-fold and 1.5-fold. This gene encodes protein pleiotropic regulatory locus 1 (Protein PRL1) regulates glucose, stress, hormone responses, and cytochrome P450 CYP90A1/CPD, and coordinates the expression of genes related to cell wall modification and growth. As a component with the MAC complex, it regulates defense responses through transcriptional control that essential for plant innate immunity. By suppressing the expression of several (1)O(2)-responsive genes, PRL1 seems to play a major role in modulating responses of plants to environmental changes by interconnecting (1)O(2)-mediated retrograde signaling with other signaling pathways. Also, PRL1 acts as inhibitor of AKIN10 and AKIN1 (Palma K et al., 2007). Gene (CF806709) encoding germin-like protein subfamily 1 member 7 that may play a role in plant defense was up-regulated by 3.5-fold at 12dpi and 3.9-fold at 8dpi. Transcript (CD398801) encoding Superkiller Viralicidic Activity 2-like that may be involved in the antiviral activity by blocking translation of poly (A) deficient mRNAs was shown to be up-regulated in both samples by 2.0 fold at 12dpi and 1.8 fold at 8dpi. This gene encodes a DEAD box protein that is implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly.

Our data reflect that numerous gene involved in the transcription regulation were shown to be differentially expressed in common at both time points. Gene (CD405069) encoding transcription factor bHLH63 was up-regulated at both time

points 12dpi and 8dpi by 2.0 fold and 1.6 fold, respectively. This protein is a sequence-specific DNA binding transcription factor that act as a positive regulator of flower development. Gene (CD398706) encoding transcription factor bHLH77 was induced at 12dpi 1.9-fold and at 8dpi 3.1-fold. The expression of gene (BF598552) encoding ethylene-responsive transcription factor ERF113 was increased by 6.7-fold at 12dpi and 1.5-fold at 8dpi. This gene acts as a transcriptional activator, binds to the GCC-box pathogenesis-related promoter element and may be involved in the regulation of gene expression by stress factors. On the other hand, two of the genes encoding transcriptional factors are shown to be suppressed at both time points. For example, the expression of gene (BQ785406) encoding trihelix transcription factor GT-2 was reduced by -1.8 fold and -1.7 fold at 12dpi and 8dpi, respectively. Study showed this gene improves plant tolerance to abiotic stresses (Xie *et al*. 2009). The transcription of gene (BE659926) encoding nuclear transcription factor Y subunit C-9 was suppressed at 12dpi by -1.8 fold and at 8dpi by -1.5 fold. This gene stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.

3.7 Conclusion

There are major changes in the gene expression between the soybean infected by SCN and the soybean infected by RKN. We identified changes in important genes and pathways during the infection particularly after 12 days for RKN infection and 8 days for SCN infection. The genes identified here represent an additional and significant pool of information to take into consideration and explore with regard to the interaction between *G.max* and *H.glycines* and also between *G.max* and *M.Incognita*. The genes can be investigated in functional analyses to provide new insights into host-parasite interactions. In the future, some of these genes may be used to control the plant parasitic infestation. For example, by over-expressing some of the genes involved in the defense system or by silencing others.

Chapter 4

Database design and development

The Root Knot Nematode Lethal Database (RKNLDB) is a public online database that links RKN ESTs with RNAi experiments and their phenotypes in *C.elegans*. The main focus for our database is the identification of the lethal genes in *Meloidogyne.spp* that can produce lethal phenotype if mutated or silenced.

RKNLDB is a relational database that was built on SQLServer2008 and it is housed at the bioinformatics server in Towson University domain, MD, USA. The database was implemented on a server running Windows 2003 server. RKNLDB was established in 2011 to serve as a sequence database for the root knot nematode community with the primary focus of store and analyze RKN ESTs, genomics, and proteins to identify our candidate lethal genes.

RKNLDB is mainly an archival and analysis database that was designed to handle the data generated by conducting a comparative genomics and protein analysis, comparing RKN genes of six species: *M. arenaria, M. chitiwoodi, M. hapla, M. incognita, M. javanica,* and *M. paranaensis* to protein of the *C. elegans* that was downloaded from WormBase server.

RKNLDB is divided into three related entities, sequence of the best hits, RNAi experiments, and all available Phenotypes. The RKN sequences were downloaded from nematode.net. The result of the BLASTX was parsed to get our best hits and then it was imported into RKNLDB database. We used one of our primary entities

(WormBase Gene ID) to link RKNLDB to Wormbase server. By submitting the available WBGeneIDs in our local database we were able to identify all the RNAi experiments and then we imported back to RKNLDB. We used WormMart tool to identify all the phenotypes for each RNAi we have in our local database then the results were imported back into RKNLDB. Elements are also blasted with each other to see if there exists similarities between them, the similar clones and their corresponding e-values are recorded.

4.1 Description of the Entities and Relationships in RKNLDB

RKNLDB stores all the best hits as results of the comparative genomics analysis, comparing RKN genes of the six species to *C. elegans* via WormBase. All the RKN EST sequences and the *C. elegans* protein sequences are stored there. In each hit there is RKN contig that has unique ID and also there is *C. elegans* protein that has unique ID. More parameters are stored in our local database: score, e-value, and identity. The entity-relation diagram for RKNLDB is depicted in figure 17.

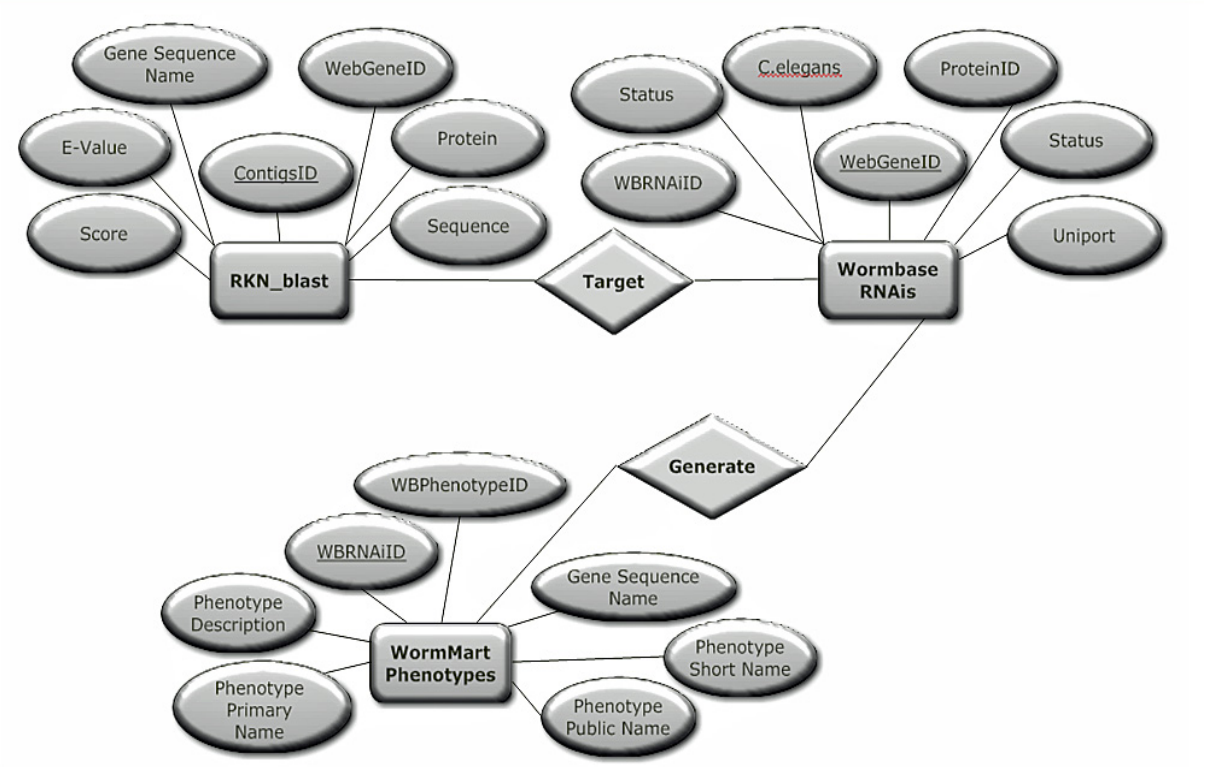


Figure 17: ER-diagram for RKNLDB Database

4.2 Database schema and Implementation

A database schema is the implementation of the ER- diagram. RKNLDB's schema conform the minimal information about RNAi experiments in *C. elegans* and all the phenotypes available for each experiment. Appendix A shows the schema for RKNLDB listing the tables and the columns associated with each, along with the data format of each column. The relationships between tables are also shown through the primary and foreign keys. A primary key is a unique identifier for every row in each table, while the foreign key is an identifier used to link rows in one table with rows in another table. For instance in the table "RKN_Wormbase" there is a primary key

WBGeneID which is also the foreign key cause it references the primary key in table "RKN_blast". RKNLDB capture a lot of the metadata associated with our experiment of identifying the candidate lethal genes in *Meloidogyne.spp.* Figure 6 shows a relational diagram for RKNLDB and it is a graphical depiction of the tables and their relationships. The SQL implementation for RKNLDB can be found in appendix B. The SQL code was written for SQL Server2008, but the code is cross platform and can be used to implement RKNLDB on any relational database management system (Oracle, Sybase or MySQL) with slight modification for the syntaxes.

4.3 RKNLDB Web Interface

RKNLDB is a relational database built on SQL Server2008 and is housed at (bioinformatics.towson.edu) Towson University in Towson, MD. The database server is housed on a dual Intel Pentium server with processor speeds of 2.80 GHz with 8 GB of memory. The web server is Microsoft IIS (Internet Information Services) version 6.0 running on a Windows 2003 server operating system. Active Server Pages (ASP.Net) was used to build the user interfaces and query pages. SQL codes are embedded within ASP.Net to query the database.

RKNLDB's web site is publicly available site (http://bioinformatics.towson.edu/RKN/). All published data and all the functionality of the site is available on the public site. The web interface to RKNLDB provides three links at the top of the home page to allow researchers to access: lethal RKN that allow end users browsing and searching RKN candidate lethal genes, Microarray

Experiments for browsing and performing queries over all soybean microarray data and sequences, and all generated charts (see Figure 18).

4.3.1 Searching lethal genes

RKNLDB web is showing one of the user interfaces to query all the candidate lethal genes over all six species of the *Meloidogyne.spp* and also user can select between them. Also user can select one of the lethal stages to be displayed. The user can search lethal genes according to their level of similarity buy choosing between the five cut offs of the E-value (see Figure 19).

Figure 18: RKNLDB's home page on the web.

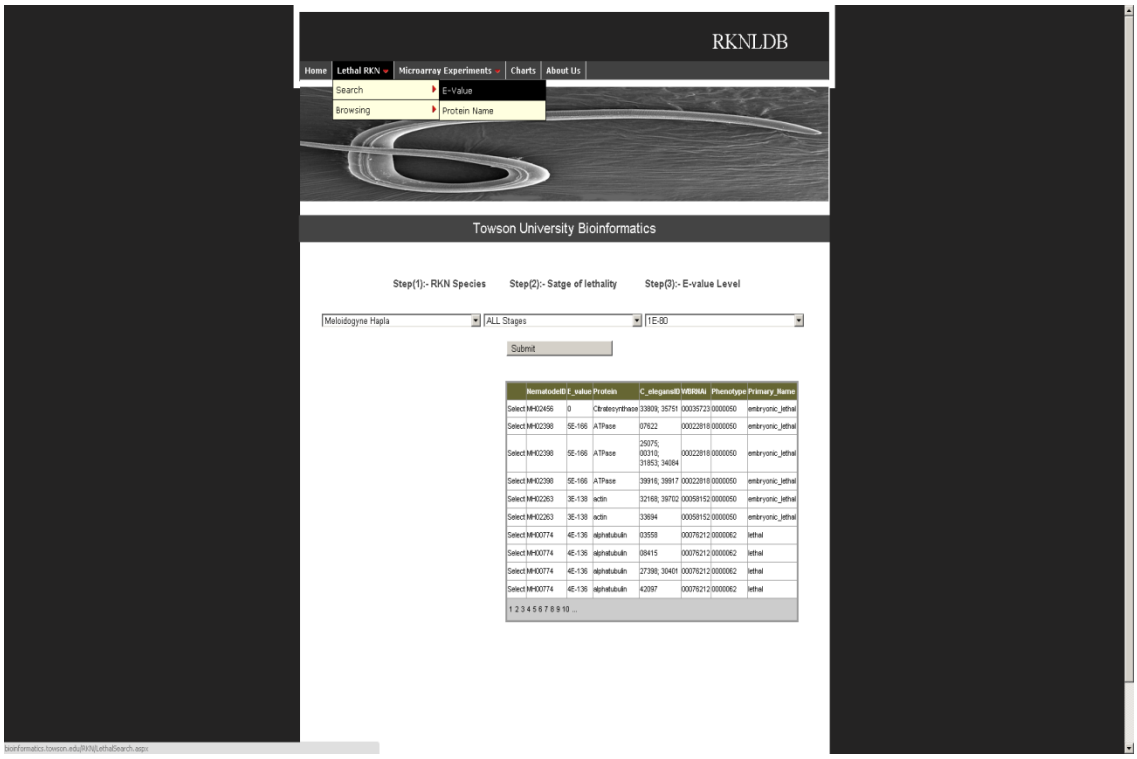


Figure 19: Screen snap shot of one of RKNLDB's user interfaces, this one showing the different types of queries available for the users to search our lethal genes database.

4.3.2 Sequence database

The sequence database provides access to EST sequences and amino acid sequences stored in our local database. By clicking on the select link the user will be able to display the nucleotide sequence and the protein sequence for the selected lethal gene (Figure 20). Among the queries available to users through the web interface is to query using individual protein name to help researcher being more specific in their search (Figure 21).

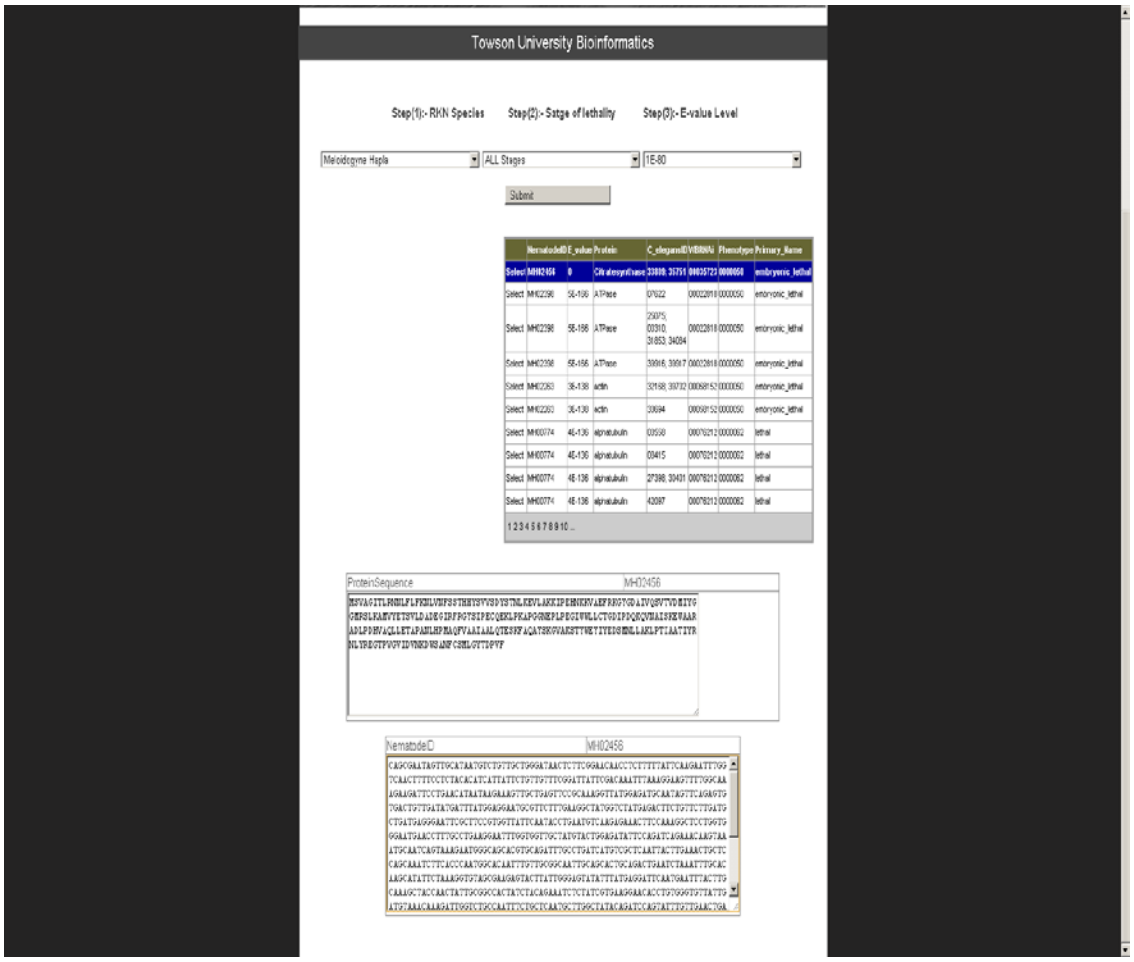


Figure 20: Screen snap shot of RKNLDB showing the selected lethal gene associated with its nucleotide and protein sequences.

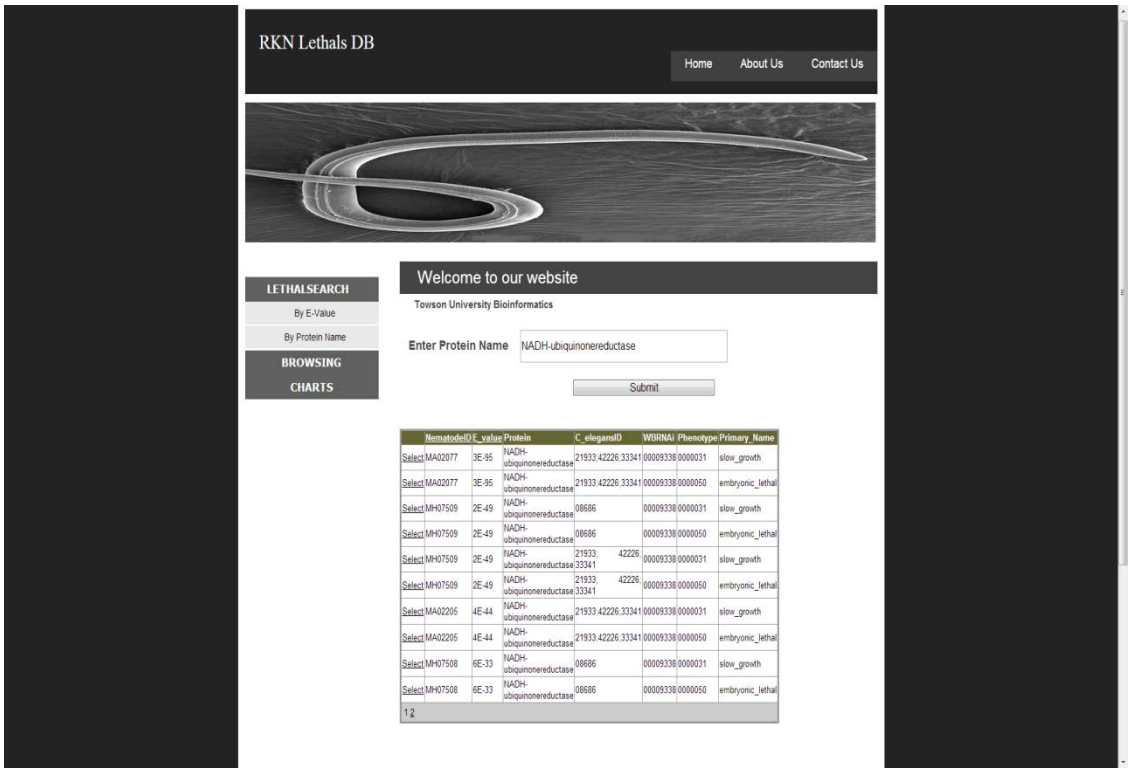


Figure 21: Screen snap shot of RKNLDB showing a user interface to query particular protein name.

	NematodeID	E_Value	ProteinID	WBRNAiID	WBPhenotypeID	Phenotype_Primary_Name
Select	MC04476	2.3		00067588	0000031	slow_growth
Select	MC04477	4.1		00051941	0000050	embryonic_lethal
Select	MC04477	4.1		00051941	0000535	organism_morphology_variant
Select	MC04477	4.1		00051941	0000689	maternal_sterile
Select	MC04477	4.1		00051941	0000886	Variant
Select	MC04478	3E-30	histoneH2A	00030041	0000049	postembryonic_development_variant
Select	MC04478	3E-30	histoneH2A	00030041	0000062	lethal
Select	MC04478	3E-30	histoneH2A	00030041	0000886	Variant
Select	MC04479	1.2	glycoprotein	00078833	0001271	pathogen_induced_death_increased
Select	MC04480	1E-96	tubulinbeta-chain	00064965	0000679	transgene_subcellular_localization_variant
1 2 3 4 5 6 7 8 9 10 ...						

Figure 22: Snap shot showing the result of a query with the gene ID and all relevant information on that gene.

4.3.3 Browsing

RKNLDB is showing another user interface to query all the data available in our local data either its lethal or not. This interface provides end users with all the similarity between each one of the six RKN species and *C.elegans* with providing the ability to display the EST and protein sequences (see figure 23).

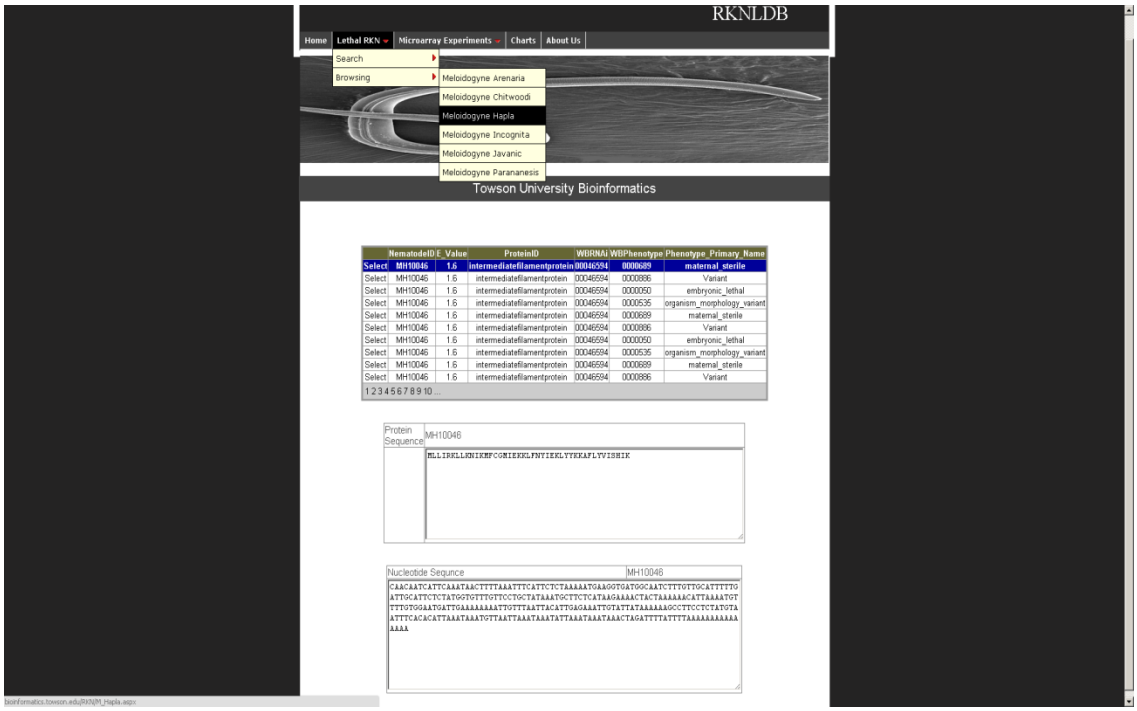


Figure 23: Screen snap shot of RKNLDB showing a user interface to brows the entire database for each species of RKN.

4.4 Soybean Microarray Database

The soybean microarray database attempts to provide an integrated view of the interaction of soybean with the SCN and RKN on different time points to study the soybean gene expression profiles. The microarray database is public on (http://bioinformatics.towson.edu/RKN/)

The primary purpose of the microarray database is to store and analyze EST, genes, and microarray data generated from experiments involving the interaction of soybean with the SCN and RKN. The database was designed to handle data generated by DNA microarray experiments. This microarray database involves two microarray

experiments, the first one was to study genes expressions profile after the infection of RKN at different time points by printing of DNA elements or “genes” onto a glass slide; the slides are then hybridized with three different cDNAs (cDNA is known as the RNA probe) each labeled with a different fluorescent dye. Three populations of RNA probes are generated; one containing the control RNA and the other two are containing the test RNA collected at two different time points 12dai and 10wai, respectively. These two “RNA Probes” are then hybridized to the DNA elements on the slide. After hybridization, a laser scanner detects the amount of the fluorescence on the slide providing raw data that is further analyzed. This process repeated three times according to having three replicates of each sample and resulting in having three replicates for each RNA label. The goal is to identify the genes whose mRNA is expressed more or less in the 12dai and 10wai labeled RNAs as compared to the control labeled one or vice versa. Similarly, second experiment was to study genes expressions profile after the infection of Cyst Nematode at 8 dpi. The database is mainly an archival/analysis database but serves as a LIMS (Laboratory information management system) database as well.

4.4.1 Database schema and Implementation

The entity relation-diagram for the microarray database is depicted in figure 6. RKNLDB stores also relevant data relating to the microarray experiments, such as raw data, normalized data, GOs, and annotation. RKNLDB stores the unique clone ID for the elements (genes) that are printed on the glass slide the DNA sequence, the name of the species it was isolated from, and a Genbank accession number if it has

one, are all stored. Elements are blast searched and for each clone ID there exits a summary of that blast search that includes: best hit, identities, percent identities, score, e-value and blast type.

Microarray database schema conforms to the minimal information about the microarray experiment. Appendix A shows the schema for the microarray database listing the tables and the columns associated with each, along with the data format of each column (i.e. whether it's an integer or string). The relationships between tables are also shown through the primary and foreign keys. The microarray database captures a lot of the metadata associated with microarray experiments. Figure 24 below shows a relational diagram for the microarray database, a graphical depiction of the tables and their relationships.

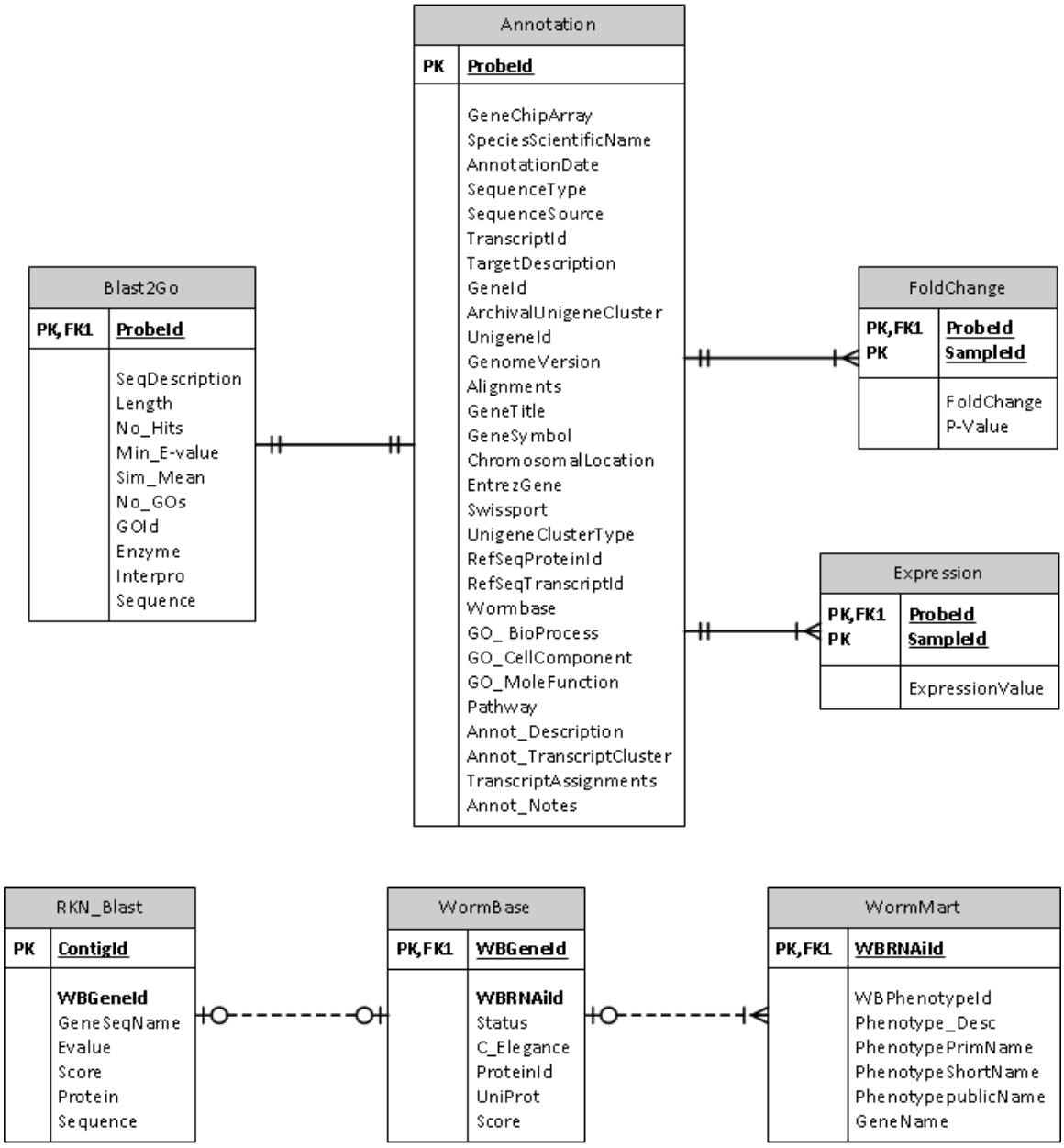


Figure 24: Relational diagram for Soybean microarray database.

4.4.2 Microarray Web Interface

Microarray database is part of our RKNLDB web interface. First page is showing a user interface to query all the data available in our local data for both RKN and SCN microarray experiments. This interface provides end users with all the information

about the differentially expressed genes (see Figure 25) with providing the ability to display all the annotation information for particular gene by clicking on it (see Figure 26).

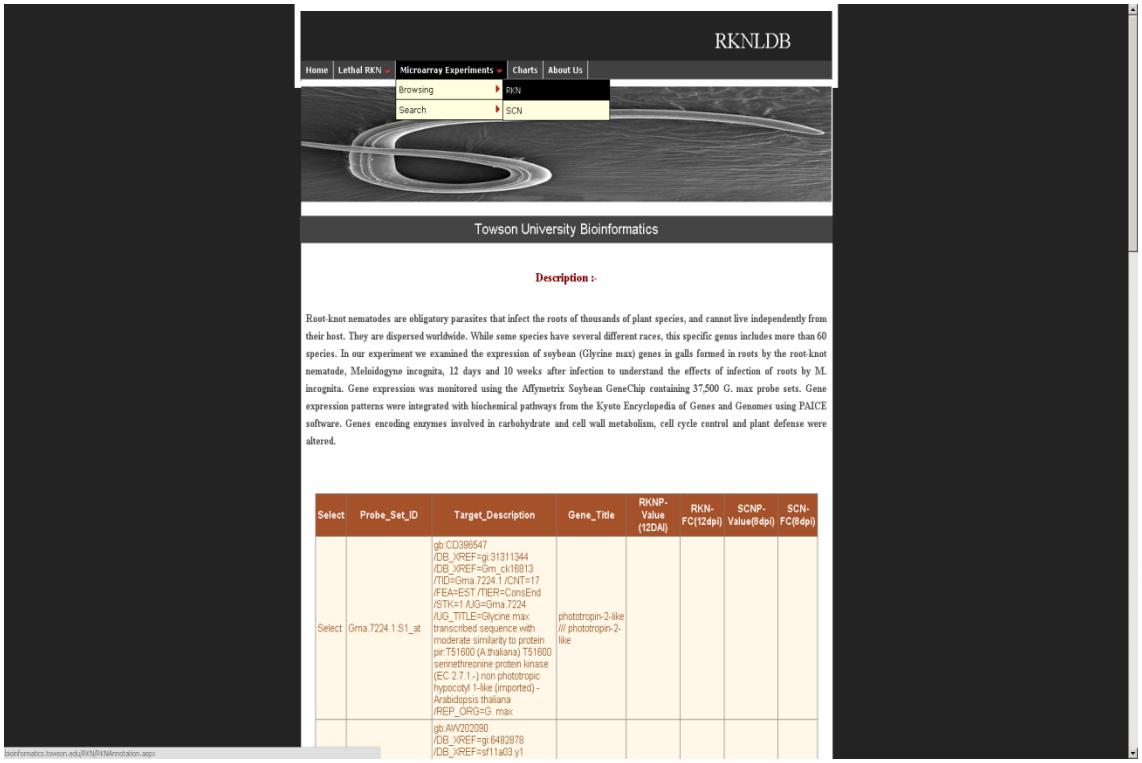


Figure 25: Screen snap shot of user interface to brows the entire microarray database for RKN and SCN.

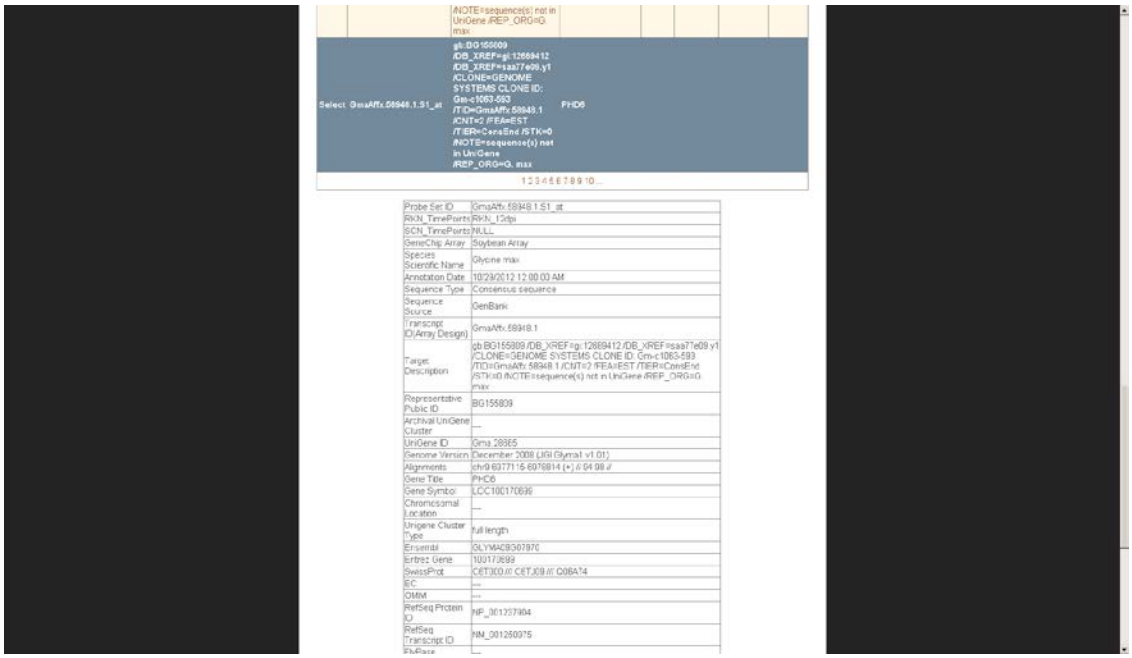


Figure 26: Snap shot to show the annotation description when the user clicks on a particular gene

Another user interfaces was added to query all the microarray database and users can search the differentially expressed genes according to the range of the fold change by choosing the minimum and maximum levels cut offs and also users can choose between only RKN, only SCN, or both experiments (see Figure 27). Users can also query the database by the gene name (see Figure 28). Finally, the users can query the database using the probe ID (see Figure 29).

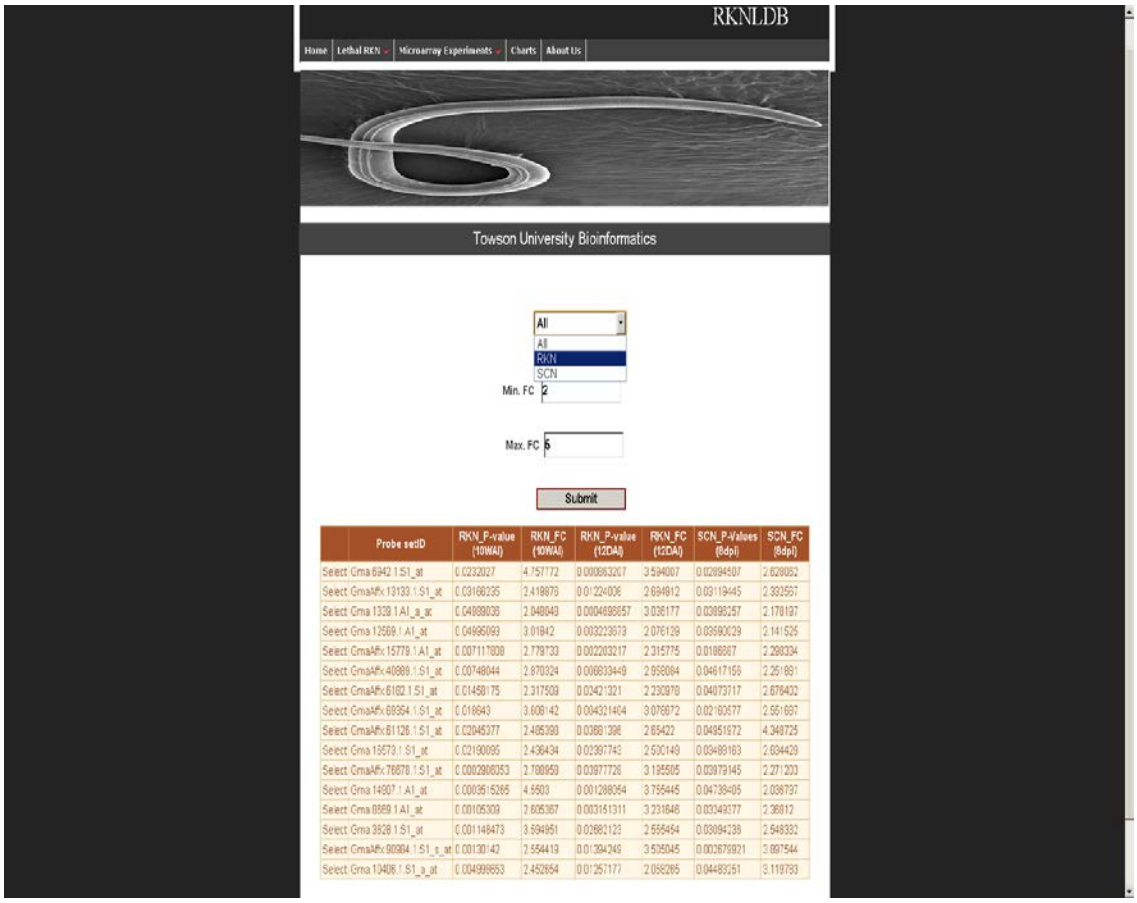


Figure 27: Screen snap shot to show queries available for the users to search our microarray database using Fold change level.

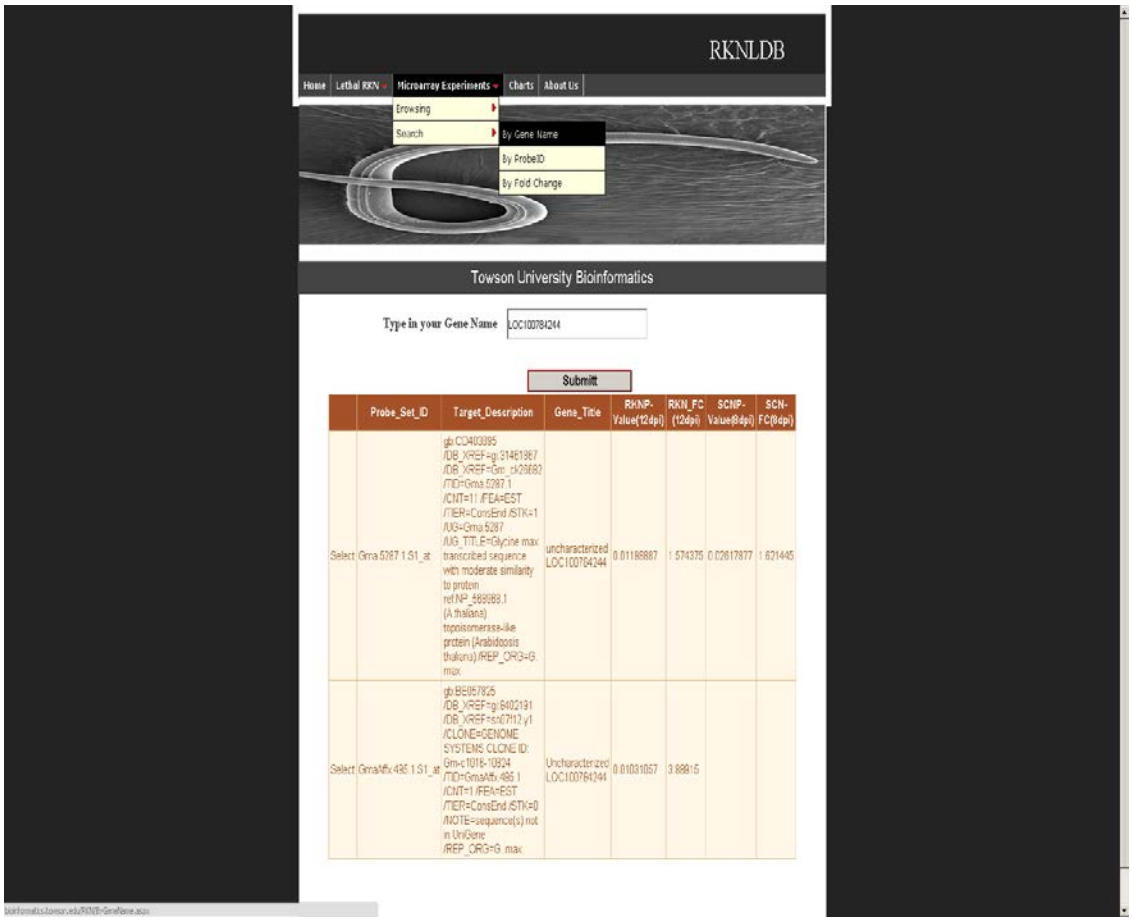


Figure 28: Screen snap shot to show queries available for the users to search our microarray database using gene name or location.

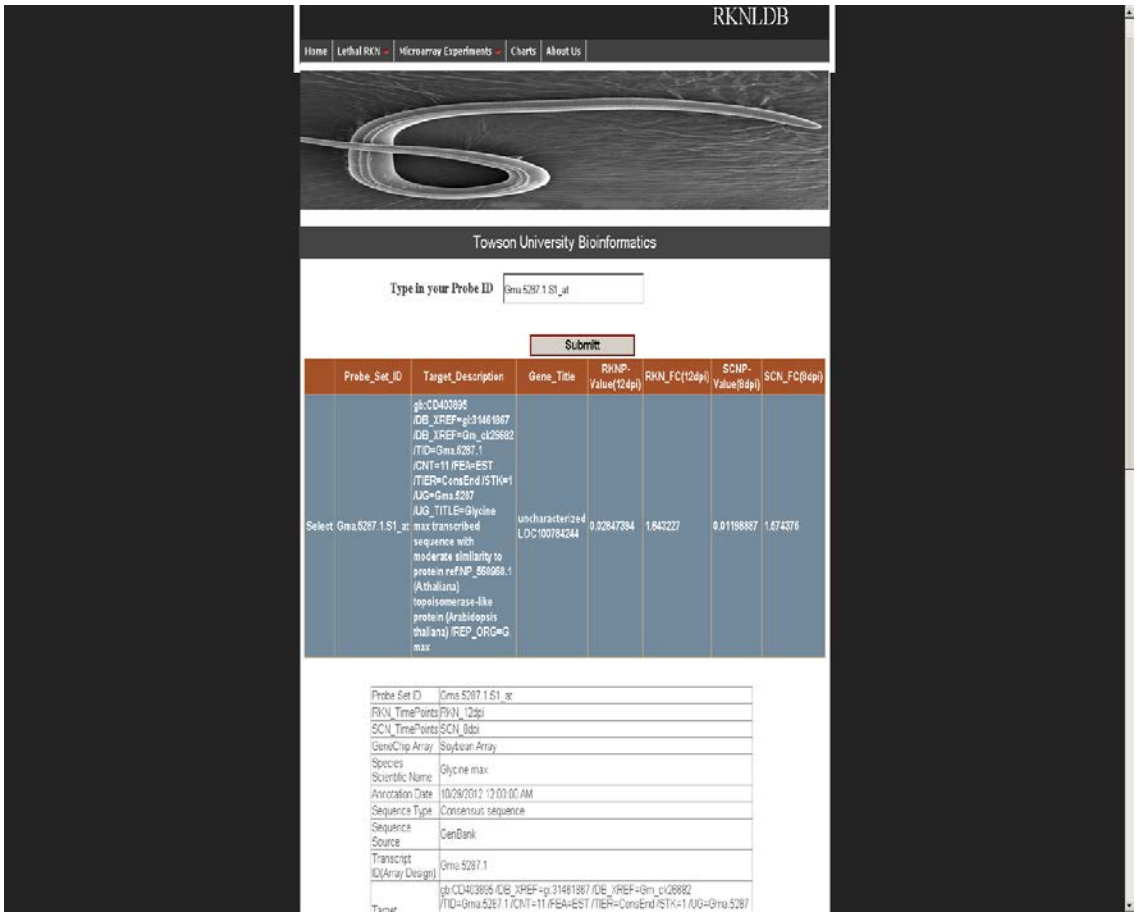


Figure 29: Screen snap shot to show queries available for the users to search our microarray database using Probe ID.

4.5 Conclusion

RKNLDB is unique among other microarray databases because it integrates many of the analysis methods for DNA microarrays within the database itself, which helps analyze and interpret data faster. RKNLDB is a customized local laboratory database; it is not a repository database. RKNLDB fulfills the roles and needs of the microarray users and researchers in soybean and the genes expression profile after the infection with RKN and SCN, as well as serve as an archival and analysis database for microarray experiments. RKNLDB is smaller in its structure and scope, and hence runs faster because there are fewer tables to join. Web based applications and tools were developed as a web interface to query the local database from the World Wide Web and to handle the data that will be generated. RKNLDB database and the web interface will serve the scientific community by providing the research community with accurate, current, accessible information concerning the genetics, genomics and biology of the soybean host-parasite interactions with the related nematodes RKN and SCN. In the coming years RKNLDB will be routinely updated to provide scientists and researchers with some information about these genes that may be used to control the plant parasitic nematode infestation and therefore to control the annual food loss.

Chapter 5

Final Remarks and Conclusions

Functional genomic is quickly emerging as an important research field. Though great advances have been made in functional genomics in the recent years, there still remains a lot of work to be done. Especially with plant genomics which has received little attention from the research community.

Through the use of computational techniques we have discovered unique genes and pathways that might be involved in the process of controlling RKN infestation. We used EST analysis as a tool for gene discovery but according to the lack of information for the RKN species. We conducted a comparative genomics analysis, comparing RKN to C. elegans. This study showed that there is significant similarity across those two different species. Our results also indicate that there are number of genes that are lethal within different species of the RKN. The application of RNA interference to genomic studies of *C. elegans* has been very successful in helping to identify genes that play an important role in the life cycle of RKN which made us able to discover which genes are lethal at which stage of development. Results of this study are promising and will become useful in the future for controlling RKN infestation, thereby saving the world more than $100 billion that are lost every year from RKN infestation. We found EST analysis to be a useful functional genomics tool to identify the candidate lethal genes in RKN.

The emphasis in the recent years has been on human and other vertebrate genomes, with little search done on plant genomes. In our study, we attempted to study global gene expression in soybean, one of the world's major crops, particularly to study the SCN and RKN interactions. In this study we analyzed the data generated by two different microarray experiments. The goal is to identify the genes whose mRNA is expressed more in SCN labeled RNA population as compared to RKN labeled one or vice versa. Numerous applications to conduct statistical analysis have been done: background correction, normalization, and T-test. Our study indicated that only few genes were altered commonly between SCN and RKN. In the future, some of these genes may be used to control the plant parasitic infestation.

We built the root-knot nematode lethal database (RKNLDB), one of the world's first lethal genes databases, to serve as a data storage and analysis platform for both RKN EST's and soybean microarray experiments. We demonstrated how a database can be an active participant in data analysis and data mining. Through the use of SQL procedures and queries embedded in the Web based user interfaces, we successfully normalized, analyzed, and mined microarray experiments for new discoveries. RKNLDB provide the research community with accurate, current, accessible information concerning the genetics, genomics and biology of the soybean host-parasite interactions.

In the near future we expect functional genomics to become an even more important field. New tools and techniques, both in molecular biology and bioinformatics will be developed to facilitate greater understanding of the SCN and RKN interaction. Once the SCN and RKN genome is fully sequenced, the use of the tools and methods

developed in this study will become even more useful. New tools for gene identification and classification, time series analysis and visualization could also be integrated into our database system for novel gene discoveries. Real time visualization of SQL queries could be especially useful as well as 3D interactive software to mine biological pathways in conjunction with microarray data.

Appendices

Appendix A:

RKNLDB Schema

Note PK stands for Primary Key and FK stands for Foreign Key.

RKN_blast (ContigsID: Inetger, Score: float, E-value: float, GeneSeqName: nvarchar, WBGeneID: Integer, ProteinName: string, Sequence: string)

PK: ContigsID **FK**: WBGeneID references Wormbase_RNAi

Wormbase_RNAi (WBGeneID: Integer, CelegnasID: Integer, ProteinID: Integer, Status: string, UniProt: nvarchar, Locus: nvarchar, WBRNAiID: Integer)

PK: WBGeneID **FK**: WBRNAiID references WormMart_Phenotypes.

WormMart_Phenotypes (WBRNAiID: Integer, WBPhenotypeID: Integer, Phenotype Primary Name: string, Phenotype Short Name: string, Phenotype Description: string, Gene Public Name: nvarchar, Gene Seq Name: nvarchar)

PK: WBRNAiID

Microarray database Schema

Expression (ProbeId: nvarchar, SampleId: integer, ExpressionValue: decimal)

PK: ProbeId, SampleId **FK**: SampleId

FoldChange (ProbeId: nvarchar, SampleId: integer, FoldChange: decimal, P-Value: decimal)

PK: ProbeId, SampleId **FK**: SampleId

Annotation (ProbeId: nvarchar, GeneChip Array: nvarchar, Species Scientific Name: nvarchar, Annotation Date: datetime, Sequence Type: nvarchar, Sequence Source: nvarchar, Transcript ID(Array Design): nvarchar, Target Description: nvarchar, GeneID: varchar, Archival UniGene Cluster: nvarchar, UniGene ID: nvarchar, Genome Version: nvarchar, Alignments: nvarchar, Gene Title: nvarchar, Gene Symbol: char, Chromosomal Location: nvarchar, Entrez Gene: integer, SwissProt: char, Unigene Cluster Type: char, RefSeq Protein ID: nvarchar, RefSeq Transcript ID: nvarchar, Gene Ontology Biological Process: nvarchar, Gene Ontology Cellular Component: nvarchar, Gene Ontology Molecular Function: nvarchar, Pathway: nvarchar, Annotation Description: nvarchar, Transcript Assignments: nvarchar, Annotation Notes: nvarchar)

PK: ProbeId

Blast2go (ProbeId: nvarchar, Sequence Description: nvarchar, length: integer, Number of hits: integer, min. e-value: nvarchar, sim mean: nvarchar, Number of GOs: integer, GOs: nvarchar, Enzyme: nvarchar, InterPro: nvarchar, Sequence: nvarchar)

PK, FK: ProbeId

Appendix B: SQL implementation of RKNLDB

```
/* RKNLDB Database Schema
   Author: Ahmed Ismail
   Department of Computer and Information Science
   Towson University
*/
/* Dropping all tables*/
Drop table RKN_blastP
Drop table Wormbase_RNAi
Drop table WormMart_Phenotypes
Drop table Expression
Drop table FoldChange
Drop table Annotation
Drop table Blast2go
/* Table: RKN_blastP */
create table RKN_blastP (
   ContigsID         Integer,
   Score:             float,
   E-value:           float,
   GeneSeqName     nvarchar,
   WBGeneID         Integer,
   ProteinName       string,
   Locus             nvarchar,
   Sequence          string,
PRIMARY KEY (ContigsID)
)
```

```
/* Table: Wormbase_RNAi */
create table Wormbase_RNAi (
  WBGeneID        Integer,
  CelegnasID      Integer,
  ProteinID       Integer,
  Status          string,
  UniProt         nvarchar,
  WBRNAiID        Integer,
PRIMARY KEY (WBGeneID)
)
/* Table: WormMart_Phenotypes */
create table WormMart_Phenotypes (
WBRNAiID                  Integer,
WBPhenotypeID             Integer,
Phenotype Primary Name    string,
Phenotype Short Name      string,
Phenotype Description     string,
Gene Public Name          nvarchar,
Gene Seq Name             nvarchar
PRIMARY KEY (WBRNAiID)
)
/* Table: Expression */
create table Expression (
ProbeId           nvarchar,
SampleId          integer,
ExpressionValue   decimal
PRIMARY KEY (ProbeSetID) )
```

/* Table: **FoldChange */**

create table FoldChange (

ProbeId nvarchar,

SampleId integer,

FoldChange decimal,

P-Value decimal

PRIMARY KEY (ProbeId, SampleId)

)

/* Table: **Annotation */**

create table Annotation (

ProbeId: nvarchar,

GeneChip Array: nvarchar,

Species Scientific Name: nvarchar,

Annotation Date: datetime,

Sequence Type: nvarchar,

Sequence Source: nvarchar,

Transcript ID(Array Design): nvarchar,

Target Description: nvarchar,

GeneID: varchar,

Archival UniGene Cluster: nvarchar,

UniGene ID: nvarchar,

Genome Version: nvarchar,

Alignments: nvarchar,

Gene Title: nvarchar,

Gene Symbol: char,

Chromosomal Location: nvarchar,

Entrez Gene: integer,

SwissProt: char,

Unigene Cluster Type: char,

RefSeq Protein ID: nvarchar,

RefSeq Transcript ID: nvarchar,

Gene Ontology Biological Process: nvarchar,

Gene Ontology Cellular Component: nvarchar,

Gene Ontology Molecular Function: nvarchar,

Pathway: nvarchar,

Annotation Description: nvarchar,

Transcript Assignments: nvarchar,

Annotation Notes: nvarchar

PRIMARY KEY (ProbeId)

)

/* Table: **Blast2go */**

create table Blast2go (

ProbeId: nvarchar,

Sequence Description: nvarchar,

length: integer,

Number of hits: integer,

min. e-value: nvarchar,

sim mean: nvarchar,

Number of GOs: integer,

GOs: nvarchar,

Enzyme: nvarchar,

InterPro: nvarchar,

Sequence: nvarchar

PRIMARY KEY (ProbeId))

Appendix C: Identified lethal genes in all six species of Meloidogyne.spp

Table C.1: Identified lethal genes in *Meloidogyne arenaria* at different stages of development.

WBGeneID	Phenotype Primary Name	Gene Public Name	Gene Seq Name
00006639	embryonic_lethal	tsp-13	Y39B6A.6
00006639	larval_lethal	tsp-13	Y39B6A.6
00015168	embryonic_lethal	tag-320	B0403.4
00015168	larval_lethal	tag-320	B0403.4
00021562	embryonic_lethal	nuo-5	Y45G12B.1
00021747	embryonic_lethal	H24K24.3	H24K24.3
00021747	embryonic_lethal	Y50D4C.2	Y50D4C.2

Table C.2: Identified lethal genes in *Meloidogyne chitiwoodi* at all stages of development.

WBGeneID	Phenotype Primary Name	Gene Public Name	Gene Seq Name
00000036	lethal	ace-2	Y44E3A.2
00000041	embryonic_lethal	aco-2	F54H12.1
00000051	embryonic_lethal	acr-12	R01E6.4
00000051	lethal	acr-12	R01E6.4
00000057	embryonic_lethal	acr-18	F28F8.1
00000057	larval_lethal	acr-18	F28F8.1
00000072	embryonic_lethal	add-1	F39C12.2

00000072	larval_lethal	add-1	F39C12.2
00000082	embryonic_lethal	adt-1	C02B4.1
00000086	embryonic_lethal	aex-3	C02H7.3
00000086	larval_lethal	aex-3	C02H7.3
00000109	embryonic_lethal	alh-3	F36H1.6
00000109	larval_lethal	alh-3	F36H1.6
00000110	embryonic_lethal	alh-4	T05H4.13
00000110	larval_lethal	alh-4	T05H4.13
00000113	embryonic_lethal	alh-7	F45H10.1
00000113	embryonic_lethal	ubc-15	Y110A2AR.2
00000113	larval_lethal	alh-7	F45H10.1
00000113	larval_lethal	ubc-15	Y110A2AR.2
00000117	lethal	alh-11	F42G9.5
00000122	embryonic_lethal	aly-3	M18.7
00000140	adult_lethal	anc-1	ZK973.6
00000140	embryonic_lethal	anc-1	ZK973.6
00000140	larval_lethal	anc-1	ZK973.6
00000140	lethal	anc-1	ZK973.6
00000157	embryonic_lethal	aps-2	F02E8.3
00000157	larval_lethal	aps-2	F02E8.3
00000165	embryonic_lethal	aps-3	Y48G8AL.14
00000171	embryonic_lethal	aqp-3	Y69E1A.7

00000171	larval_lethal	aqp-3	Y69E1A.7
00000175	embryonic_lethal	aqp-7	M02F4.8
00000188	embryonic_lethal	arl-3	F19H8.3
00000188	embryonic_lethal	tps-2	F19H8.1
00000188	larval_lethal	arl-3	F19H8.3
00000188	larval_lethal	tps-2	F19H8.1
00000197	embryonic_lethal	aars-2	F28H1.3
00000200	embryonic_lethal	arx-2	K07C5.1
00000224	embryonic_lethal	atgp-1	F26D10.9
00000224	larval_lethal	atgp-1	F26D10.9
00000226	embryonic_lethal	atl-1	T06E4.3
00000226	embryonic_lethal	col-146	T06E4.6
00000226	embryonic_lethal	col-147	T06E4.4
00000226	embryonic_lethal	nspa-11	T06E4.13
00000226	embryonic_lethal	T06E4.5	T06E4.5
00000227	embryonic_lethal	atm-1	Y48G1BL.2
00000227	larval_lethal	atm-1	Y48G1BL.2
00000240	lethal	pah-1	K08F8.4
00000277	lethal	cab-1	C23H4.1
00000281	embryonic_lethal	cah-3	K05G3.3
00000281	larval_lethal	cah-3	K05G3.3
00000285	embryonic_lethal	cal-1	C13C12.1

00000285	larval_lethal	cal-1	C13C12.1
00000369	embryonic_lethal	ccf-1	Y56A3A.20
00000369	larval_lethal	ccf-1	Y56A3A.20
00000376	embryonic_lethal	ccr-4	ZC518.3
00000378	lethal	cct-2	T21B10.7
00000395	embryonic_lethal	cdh-3	ZK112.7
00000395	larval_lethal	cdh-3	ZK112.7
00000397	embryonic_lethal	cdh-5	F08B4.2
00000397	larval_lethal	cdh-5	F08B4.2
00000403	lethal	casy-1	B0034.3
00000407	lethal	cdk-5	T27E9.3
00000421	embryonic_lethal	ced-7	C48B4.4
00000426	embryonic_lethal	ced-12	Y106G6E.5
00000445	embryonic_lethal	ceh-22	F29F11.5
00000445	larval_lethal	ceh-22	F29F11.5
00000454	embryonic_lethal	ceh-33	C10G8.7
00000481	embryonic_lethal	cha-1	ZC416.8
00000481	lethal	cha-1	ZC416.8
00000500	larval_lethal	chn-1	T09B4.10
00000502	embryonic_lethal	chp-1	Y110A7A.13
00000506	embryonic_lethal	cic-1	H14E04.5
00000510	embryonic_lethal	cka-2	C52B9.1

00000510	larval_lethal	cka-2	C52B9.1
00000515	embryonic_lethal	ckc-1	T27A10.3
00000515	larval_lethal	ckc-1	T27A10.3
00000525	embryonic_lethal	clc-4	T05A10.2
00000525	larval_lethal	clc-4	T05A10.2
00000533	embryonic_lethal	clh-6	R07B7.1
00000540	embryonic_lethal	cln-3.2	C01G8.2
00000542	embryonic_lethal	clp-1	C06G4.2
00000553	embryonic_lethal	cmk-1	K07A9.2
00000553	larval_lethal	cmk-1	K07A9.2
00000554	embryonic_lethal	cnb-1	F55C10.1
00000623	embryonic_lethal	col-46	Y18H1A.13
00000623	larval_lethal	col-46	Y18H1A.13
00000649	embryonic_lethal	col-17	F11G11.10
00000649	embryonic_lethal	col-73	F11G11.12
00000678	embryonic_lethal	col-104	F58F6.1
00000678	embryonic_lethal	F58F6.t1	F58F6.t1
00000678	larval_lethal	col-104	F58F6.1
00000678	larval_lethal	F58F6.t1	F58F6.t1
00000692	embryonic_lethal	col-118	T11B7.3
00000722	embryonic_lethal	col-149	B0024.1
00000737	embryonic_lethal	col-164	T21D9.1

00000737	larval_lethal	col-164	T21D9.1
00000738	embryonic_lethal	col-165	F14H12.1
00000738	embryonic_lethal	dpy-14	H27M09.4
00000753	early_larval_lethal	col-180	C44C10.1
00000753	embryonic_lethal	col-180	C44C10.1

Table C.3: Identified lethal genes in *Meloidogyne Hapla* at all stages of development.

WBGeneID	Phenotype Primary Name	Gene Public Name	Gene Seq Name
00000001	embryonic_lethal	aap-1	Y110A7A.10
00000007	lethal	aat-6	T11F9.4
00000020	embryonic_lethal	abt-2	F12B6.1
00000020	larval_lethal	abt-2	F12B6.1
00000022	embryonic_lethal	abt-4	Y39D8C.1
00000022	embryonic_lethal	Y39D8C.2	Y39D8C.2
00000023	embryonic_lethal	abt-5	Y53C10A.9
00000036	lethal	ace-2	Y44E3A.2
00000038	embryonic_lethal	ace-4	Y48B6A.7
00000038	larval_lethal	ace-4	Y48B6A.7
00000039	larval_lethal	acn-1	C42D8.5
00000043	embryonic_lethal	acr-3	K11G12.7
00000043	larval_lethal	acr-3	K11G12.7
00000054	embryonic_lethal	acr-15	F25G6.4

00000054	larval_lethal	acr-15	F25G6.4
00000056	lethal	21ur-14138	F53E10.7
00000056	lethal	acr-17	F53E10.2
00000057	embryonic_lethal	acr-18	F28F8.1
00000057	larval_lethal	acr-18	F28F8.1
00000071	embryonic_lethal	acy-4	T01C2.1
00000071	embryonic_lethal	M03E7.1	M03E7.1
00000072	embryonic_lethal	add-1	F39C12.2
00000072	larval_lethal	add-1	F39C12.2
00000073	embryonic_lethal	add-2	F57F5.4
00000075	embryonic_lethal	adm-4	ZK154.7
00000075	larval_lethal	adm-4	ZK154.7
00000080	lethal	adr-2	T20H4.4
00000086	embryonic_lethal	aex-3	C02H7.3
00000086	larval_lethal	aex-3	C02H7.3
00000098	embryonic_lethal	air-1	K07C11.2
00000099	embryonic_lethal	air-2	B0207.4
00000101	embryonic_lethal	aka-1	D1022.7
00000101	larval_lethal	aka-1	D1022.7
00000107	embryonic_lethal	alh-1	F54D8.3
00000109	embryonic_lethal	alh-3	F36H1.6
00000109	larval_lethal	alh-3	F36H1.6

00000111	lethal	alh-5	T08B1.3
00000114	lethal	alh-8	F13D12.4
00000118	embryonic_lethal	alh-12	Y69F12A.2
00000118	larval_lethal	alh-12	Y69F12A.2
00000120	embryonic_lethal	aly-1	C01F6.5
00000120	larval_lethal	aly-1	C01F6.5
00000137	embryonic_lethal	amx-1	R13G10.2
00000137	larval_lethal	amx-1	R13G10.2
00000138	embryonic_lethal	amx-2	B0019.1
00000138	larval_lethal	amx-2	B0019.1
00000139	embryonic_lethal	amx-3	F25C8.2
00000142	embryonic_lethal	aos-1	C08B6.9
00000155	lethal	app-1	W03G9.4
00000160	embryonic_lethal	apb-1	Y71H2B.10
00000164	lethal	apm-3	F53H8.1
00000176	embryonic_lethal	aqp-8	K02G10.7
00000180	embryonic_lethal	arc-1	ZK1320.6
00000181	lethal	ard-1	F01G4.2
00000192	embryonic_lethal	arl-8	Y57G11C.13
00000197	embryonic_lethal	aars-2	F28H1.3
00000199	embryonic_lethal	arx-1	Y71F9AL.16
00000199	embryonic_lethal	Y71F9AL.6	Y71F9AL.6

00000200	embryonic_lethal	arx-2	K07C5.1
00000201	embryonic_lethal	arx-3	Y79H2A.6
00000202	late_larval_lethal	exos-4.2	Y6D11A.1
00000218	embryonic_lethal	asp-5	F21F8.3
00000220	embryonic_lethal	atf-2	K08F8.2
00000220	larval_lethal	atf-2	K08F8.2
00000227	embryonic_lethal	atm-1	Y48G1BL.2
00000227	larval_lethal	atm-1	Y48G1BL.2
00000232	embryonic_lethal	avr-14	B0207.12
00000237	embryonic_lethal	bam-2	Y71F9AR.1
00000237	larval_lethal	bam-2	Y71F9AR.1
00000240	lethal	pah-1	K08F8.4
00000241	embryonic_lethal	bbs-1	Y105E8A.5
00000246	embryonic_lethal	bcc-1	M7.3
00000254	early_larval_lethal	bli-4	K04F10.4
00000254	embryonic_lethal	bli-4	K04F10.4
00000266	embryonic_lethal	bre-1	C53B4.7
00000275	embryonic_lethal	bub-1	R06C7.8
00000275	larval_lethal	bub-1	R06C7.8
00000276	early_larval_lethal	byn-1	F57B9.5
00000276	embryonic_lethal	byn-1	F57B9.5

Table C.4: Identified lethal genes in *Meloidogyne Incognita* at all stages of development.

WBGeneID	Phenotype Primary Name	Gene Public Name	Gene Seq Name
00000006	embryonic_lethal	aat-5	C55C2.5
00000006	larval_lethal	aat-5	C55C2.5
00000020	embryonic_lethal	abt-2	F12B6.1
00000020	larval_lethal	abt-2	F12B6.1
00000022	embryonic_lethal	abt-4	Y39D8C.1
00000022	embryonic_lethal	Y39D8C.2	Y39D8C.2
00000036	lethal	ace-2	Y44E3A.2
00000039	larval_lethal	acn-1	C42D8.5
00000041	embryonic_lethal	aco-2	F54H12.1
00000055	lethal	acr-16	F25G6.3
00000059	embryonic_lethal	acr-20	R06A4.10
00000059	larval_lethal	acr-20	R06A4.10
00000060	embryonic_lethal	acr-21	F27B3.2
00000072	embryonic_lethal	add-1	F39C12.2
00000072	larval_lethal	add-1	F39C12.2
00000073	embryonic_lethal	add-2	F57F5.4
00000074	embryonic_lethal	adm-2	C04A11.4
00000075	embryonic_lethal	adm-4	ZK154.7
00000075	larval_lethal	adm-4	ZK154.7

00000079	embryonic_lethal	adr-1	H15N14.1
00000079	embryonic_lethal	D2005.1	D2005.1
00000079	larval_lethal	adr-1	H15N14.1
00000079	larval_lethal	D2005.1	D2005.1
00000082	embryonic_lethal	adt-1	C02B4.1
00000089	embryonic_lethal	aex-6	Y87G2A.4
00000089	larval_lethal	aex-6	Y87G2A.4
00000098	embryonic_lethal	air-1	K07C11.2
00000099	embryonic_lethal	air-2	B0207.4
00000107	embryonic_lethal	alh-1	F54D8.3
00000108	embryonic_lethal	alh-2	K04F1.15
00000108	larval_lethal	alh-2	K04F1.15
00000109	embryonic_lethal	alh-3	F36H1.6
00000109	larval_lethal	alh-3	F36H1.6
00000110	embryonic_lethal	alh-4	T05H4.13
00000110	larval_lethal	alh-4	T05H4.13
00000111	lethal	alh-5	T08B1.3
00000113	embryonic_lethal	alh-7	F45H10.1
00000113	embryonic_lethal	ubc-15	Y110A2AR.2
00000113	larval_lethal	alh-7	F45H10.1
00000113	larval_lethal	ubc-15	Y110A2AR.2
00000114	lethal	alh-8	F13D12.4

00000115	embryonic_lethal	alh-9	F01F1.6
00000115	larval_lethal	alh-9	F01F1.6
00000117	lethal	alh-11	F42G9.5
00000134	embryonic_lethal	amt-2	F49E11.3
00000135	embryonic_lethal	amt-3	M195.3
00000140	adult_lethal	anc-1	ZK973.6
00000140	embryonic_lethal	anc-1	ZK973.6
00000140	larval_lethal	anc-1	ZK973.6
00000140	lethal	anc-1	ZK973.6
00000160	embryonic_lethal	apb-1	Y71H2B.10
00000162	lethal	apd-3	W09G10.4
00000163	embryonic_lethal	apb-3	R11A5.1
00000169	embryonic_lethal	aqp-1	F32A5.5
00000175	embryonic_lethal	aqp-7	M02F4.8
00000176	embryonic_lethal	aqp-8	K02G10.7
00000181	lethal	ard-1	F01G4.2
00000184	lethal	arf-6	Y116A8C.12
00000186	embryonic_lethal	ark-1	C01C7.1
00000187	embryonic_lethal	arl-1	F54C9.10
00000192	embryonic_lethal	arl-8	Y57G11C.13
00000198	embryonic_lethal	art-1	C15F1.6
00000198	larval_lethal	art-1	C15F1.6

00000200	embryonic_lethal	arx-2	K07C5.1
00000201	embryonic_lethal	arx-3	Y79H2A.6
00000203	lethal	arx-5	Y37D8A.1
00000206	embryonic_lethal	asb-1	F35G12.10
00000224	embryonic_lethal	atgp-1	F26D10.9
00000224	larval_lethal	atgp-1	F26D10.9
00000232	embryonic_lethal	avr-14	B0207.12
00000236	lethal	bag-1	F57B10.11
00000240	lethal	pah-1	K08F8.4
00000242	lethal	bbs-2	F20D12.3
00000254	early_larval_lethal	bli-4	K04F10.4
00000254	embryonic_lethal	bli-4	K04F10.4
00000259	embryonic_lethal	bpl-1	F13H8.10
00000266	embryonic_lethal	bre-1	C53B4.7
00000268	lethal	bre-3	B0464.4
00000274	embryonic_lethal	btf-1	F15D4.1
00000274	larval_lethal	btf-1	F15D4.1

Table C.5: Identified lethal genes in *Meloidogyne Javanica* at all stages of development.

WBGeneID	Phenotype Primary Name	Gene Public Name	Gene Seq Name
00000002	embryonic_lethal	aat-1	F27C8.1
00000020	embryonic_lethal	abt-2	F12B6.1
00000020	larval_lethal	abt-2	F12B6.1
00000022	embryonic_lethal	abt-4	Y39D8C.1
00000022	embryonic_lethal	Y39D8C.2	Y39D8C.2
00000039	larval_lethal	acn-1	C42D8.5
00000054	embryonic_lethal	acr-15	F25G6.4
00000054	larval_lethal	acr-15	F25G6.4
00000074	embryonic_lethal	adm-2	C04A11.4
00000086	embryonic_lethal	aex-3	C02H7.3
00000086	larval_lethal	aex-3	C02H7.3
00000099	embryonic_lethal	air-2	B0207.4
00000100	embryonic_lethal	ajm-1	C25A11.4
00000100	larval_lethal	ajm-1	C25A11.4
00000107	embryonic_lethal	alh-1	F54D8.3
00000108	embryonic_lethal	alh-2	K04F1.15
00000108	larval_lethal	alh-2	K04F1.15
00000110	embryonic_lethal	alh-4	T05H4.13
00000110	larval_lethal	alh-4	T05H4.13

00000111	lethal	alh-5	T08B1.3
00000122	embryonic_lethal	aly-3	M18.7
00000140	adult_lethal	anc-1	ZK973.6
00000140	embryonic_lethal	anc-1	ZK973.6
00000140	larval_lethal	anc-1	ZK973.6
00000140	lethal	anc-1	ZK973.6
00000157	embryonic_lethal	aps-2	F02E8.3
00000157	larval_lethal	aps-2	F02E8.3
00000160	embryonic_lethal	apb-1	Y71H2B.10
00000164	lethal	apm-3	F53H8.1
00000165	embryonic_lethal	aps-3	Y48G8AL.14
00000171	embryonic_lethal	aqp-3	Y69E1A.7
00000171	larval_lethal	aqp-3	Y69E1A.7
00000176	embryonic_lethal	aqp-8	K02G10.7
00000198	embryonic_lethal	art-1	C15F1.6
00000198	larval_lethal	art-1	C15F1.6
00000199	embryonic_lethal	arx-1	Y71F9AL.16
00000199	embryonic_lethal	Y71F9AL.6	Y71F9AL.6
00000236	lethal	bag-1	F57B10.11
00000237	embryonic_lethal	bam-2	Y71F9AR.1
00000237	larval_lethal	bam-2	Y71F9AR.1
00000249	embryonic_lethal	bir-1	T27F2.3

00000274	embryonic_lethal	btf-1	F15D4.1
00000274	larval_lethal	btf-1	F15D4.1
00000275	embryonic_lethal	bub-1	R06C7.8
00000275	larval_lethal	bub-1	R06C7.8
00000277	lethal	cab-1	C23H4.1
00000280	lethal	cah-2	D1022.8
00000283	embryonic_lethal	cah-5	R173.1
00000294	embryonic_lethal	cas-1	F41G4.2
00000367	embryonic_lethal	cca-1	C54D2.5
00000370	embryonic_lethal	21ur-10556	Y73F8A.602
00000370	embryonic_lethal	21ur-10676	Y73F8A.841
00000370	embryonic_lethal	21ur-11573	Y73F8A.609
00000370	embryonic_lethal	21ur-11797	Y73F8A.707
00000370	embryonic_lethal	21ur-1263	Y73F8A.145
00000370	embryonic_lethal	21ur-13350	Y73F8A.882
00000370	embryonic_lethal	21ur-1350	Y73F8A.369
00000370	embryonic_lethal	21ur-9523	Y73F8A.847
00000370	embryonic_lethal	ccg-1	Y73F8A.6
00000370	larval_lethal	21ur-10556	Y73F8A.602
00000370	larval_lethal	21ur-10676	Y73F8A.841
00000370	larval_lethal	21ur-11573	Y73F8A.609
00000370	larval_lethal	21ur-11797	Y73F8A.707

00000370	larval_lethal	21ur-1263	Y73F8A.145
00000370	larval_lethal	21ur-13350	Y73F8A.882
00000370	larval_lethal	21ur-1350	Y73F8A.369
00000370	larval_lethal	21ur-9523	Y73F8A.847
00000370	larval_lethal	ccg-1	Y73F8A.6
00000381	embryonic_lethal	cct-6	F01F1.8
00000390	embryonic_lethal	cdc-42	R07G3.1
00000395	embryonic_lethal	cdh-3	ZK112.7
00000395	larval_lethal	cdh-3	ZK112.7
00000396	embryonic_lethal	cdh-4	F25F2.2
00000421	embryonic_lethal	ced-7	C48B4.4
00000443	embryonic_lethal	ceh-20	F31E3.1
00000443	embryonic_lethal	F31E3.2	F31E3.2
00000443	lethal	ceh-20	F31E3.1
00000443	lethal	F31E3.2	F31E3.2
00000462	embryonic_lethal	ceh-41	T26C11.5
00000464	embryonic_lethal	ceh-44	Y54F10AM.4

Table C.6: Identified lethal genes in *Meloidogyne Paranaensis* at all stages of development.

WBGeneID	Phenotype Primary Name	Gene Public Name	Gene Seq Name
00000020	embryonic_lethal	aap-1	Y110A7A.10
00000020	embryonic_lethal	Y110A7A.21	Y110A7A.21
00000040	embryonic_lethal	abt-2	F12B6.1
00000040	larval_lethal	abt-2	F12B6.1
00000063	embryonic_lethal	acr-12	R01E6.4
00000063	lethal	acr-12	R01E6.4
00000080	lethal	adr-2	T20H4.4
00000092	embryonic_lethal	adt-1	C02B4.1
00000123	lethal	alh-8	F13D12.4
00000158	embryonic_lethal	aps-2	F02E8.3
00000158	larval_lethal	aps-2	F02E8.3
00000165	embryonic_lethal	apb-1	Y71H2B.10
00000165	embryonic_lethal	aps-3	Y48G8AL.14
00000193	embryonic_lethal	aqp-3	Y69E1A.7
00000193	larval_lethal	aqp-3	Y69E1A.7
00000217	lethal	arx-5	Y37D8A.1
00000229	embryonic_lethal	atl-1	T06E4.3
00000229	embryonic_lethal	col-146	T06E4.6
00000229	embryonic_lethal	col-147	T06E4.4

00000229	embryonic_lethal	nspa-11	T06E4.13
00000229	embryonic_lethal	T06E4.5	T06E4.5
00000370	embryonic_lethal	cav-2	C56A3.7
00000372	embryonic_lethal	21ur-10556	Y73F8A.602
00000372	embryonic_lethal	21ur-10676	Y73F8A.841
00000372	embryonic_lethal	21ur-11573	Y73F8A.609
00000372	embryonic_lethal	21ur-11797	Y73F8A.707
00000372	embryonic_lethal	21ur-1263	Y73F8A.145
00000372	embryonic_lethal	21ur-13350	Y73F8A.882
00000372	embryonic_lethal	21ur-1350	Y73F8A.369
00000372	embryonic_lethal	21ur-9523	Y73F8A.847
00000372	embryonic_lethal	ccg-1	Y73F8A.6
00000372	larval_lethal	21ur-10556	Y73F8A.602
00000372	larval_lethal	21ur-10676	Y73F8A.841
00000372	larval_lethal	21ur-11573	Y73F8A.609
00000372	larval_lethal	21ur-11797	Y73F8A.707
00000372	larval_lethal	21ur-1263	Y73F8A.145
00000372	larval_lethal	21ur-13350	Y73F8A.882
00000372	larval_lethal	21ur-1350	Y73F8A.369
00000372	larval_lethal	21ur-9523	Y73F8A.847
00000372	larval_lethal	ccg-1	Y73F8A.6
00000377	lethal	cyp-13A7	T10B9.10

00000379	lethal	cct-2	T21B10.7
00000397	embryonic_lethal	cdh-3	ZK112.7
00000397	embryonic_lethal	cdh-4	F25F2.2
00000397	larval_lethal	cdh-3	ZK112.7
00000410	embryonic_lethal	cdh-5	F08B4.2
00000410	larval_lethal	cdh-5	F08B4.2
00000416	embryonic_lethal	cdh-5	F08B4.2
00000416	larval_lethal	cdh-5	F08B4.2
00000421	embryonic_lethal	ced-2	Y41D4B.13
00000421	embryonic_lethal	ced-7	C48B4.4
00000421	embryonic_lethal	set-23	Y41D4B.12
00000421	larval_lethal	ced-2	Y41D4B.13
00000421	larval_lethal	set-23	Y41D4B.12
00000482	embryonic_lethal	ced-7	C48B4.4
00000485	embryonic_lethal	ced-7	C48B4.4
00000543	embryonic_lethal	cku-70	Y47D3A.4
00000543	larval_lethal	cku-70	Y47D3A.4
00000554	embryonic_lethal	clh-6	R07B7.1
00000564	embryonic_lethal	clu-1	F55H2.6
00000591	embryonic_lethal	cnb-1	F55C10.1
00000591	embryonic_lethal	cng-3	F38E11.12
00000591	embryonic_lethal	wrt-3	F38E11.7

00000612	embryonic_lethal	cnk-1	R01H10.8
00000612	larval_lethal	cnk-1	R01H10.8
00000645	lethal	col-35	C15A11.1
00000645	lethal	unc-15	F07A5.7
00000653	embryonic_lethal	col-36	C27H5.5
00000685	lethal	col-69	K01A2.7
00000699	larval_lethal	col-84	K12D12.3
00000699	lethal	col-77	M195.1
00000704	embryonic_lethal	col-111	F29B9.9
00000730	lethal	col-130	F08G5.4
00000737	embryonic_lethal	col-137	Y51H4A.9
00000737	embryonic_lethal	col-149	B0024.1
00000744	embryonic_lethal	col-154	F55C10.2
00000744	embryonic_lethal	col-155	F55C10.3
00000744	embryonic_lethal	col-157	T11F9.9
00000744	larval_lethal	col-154	F55C10.2
00000744	larval_lethal	col-155	F55C10.3

Appendix D: Perl scripts

Perl scripts that were written to parse and reformate our data before importing it into the SQL database.

- **Parsing the BLASTP, X, and N best hits**

```
#!/usr/bin/perl -w

# blast_text_parse.pl

#!/usr/bin/perl

# File: blastoutput

# read in and print out best_hits records

use strict;

use warnings;

my $query='';

my $besthit='';

my $score='';

my $E_value='';

if (not $ARGV[0]) {

die "usage: MA_blastp_output.txt\n";

}

open my $fh, $ARGV[0] or die "Can't open file $ARGV[0]";

while ( my $record = get_bs_record($fh) ) {

   ($query)= $record =~/Query= (.+)putative/;

   ($besthit) = $record =~/(>.+?)Length/s;

   $besthit =~ s/\s//g;

   ($score) = $record =~/Score = (.+?)bits/;

   $score =~ s/\s//g;

   ($E_value) = $record =~/Expect = (.+?)Method/;

   $E_value =~ s/\s+$//g;
```

```
        open (MYFILE ,">> MA_blastp_besthits.txt");
        print MYFILE "$query\t $besthit\t $score\t $E_value\n";
    }
    print "***DONE***";
    exit;
    sub get_bs_record {
    my ( $fh ) = @_;
    my $record='';
    my $saved_separator ='';
    $/ = 'Effective search space used:';
    $record = <$fh>;
    $/ = $saved_separator;
    return $record;
    }
```

- **Reformatting our data so we can import it to our SQL database.**

```
    #!/usr/bin/perl –w
    my $ID='';
    my $CE='';
    my $WBGene='';
    my $status='';
    my $protein_id='';
    my $UniProt='';
    my $locus='';
    my $description='';
    $filname = 'C:\AHMED ISMAIL\Downloads\New Text Document.txt';
    open(DATAFILE, "$filname") || die("Can't open $filname:!\n");
```

```
while (<DATAFILE>)
{
## $_ is a special Perl variable containing the current line (in this case)
## Use split() to get the fields. "\t" is the tab character.
chomp ($_);
($ID,$CE,$WBGene,$status,$protein_id,$UniProt,$locus,$description) = split('_', $_);
open (MYFILE ,">> MC_RNAi&phenotype.txt");
   print MYFILE "$ID\t $CE\t $WBGene\t $status\t $protein_id\t $UniProt\t $locus\t $description\n";
}
```

- **Perl script to extract and format all identification numbers of RNAi experiments**

```
#!/usr/bin/perl -w
# RNAi_text_parse.pl
#!/usr/bin/perl
# File: RNAi_phenotype
# read in and print out RNAi records
use strict;
use warnings;
my $WBRNAi=";
if (not $ARGV[0]) {
die "usage:RNAi_phenotype.txt\n";
}
open my $fh, $ARGV[0] or die "Can't open file $ARGV[0]";
while ( my $record = get_bs_record($fh) ) {
   ($WBRNAi)= $record =~/WBRNAi (.+)-/;
   open (MYFILE ,">> MC_RNAi.txt");
```

```
    print MYFILE "$WBRNAi\n";
}
print "***DONE***";
exit;
sub get_bs_record {
my ( $fh ) = @_;
my $record='';
my $saved_separator ='';
$/ = ']';
$record = <$fh>;
$/ = $saved_separator;
return $record;
}
```

- **Perl script to Create a tab free file**

```
$Inputfilname = 'C:\AIG\MA_DNA_LETHAL.txt';                    # Text file contains the sequences tab delimited

 open(INFILE, "$Inputfilname") || die("Can't open $Inputfilname:!\n");
 $outputfile = 'C:\AIG\MA_DNA_LETHAL_Tab_Free.fasta';
       open (DATA, ">>$outputfile");
$counter =0;
while(<INFILE>)
 {
    if($_ =~ /(.*)\t(.*)/){$counter++;print DATA "\n>$1\n$2";}
 }
  print "Done !!\n\n Number of Sequences= $counter  \n\n";
```

Appendix E: SQL queries

SQL queries were written and executed against RKNLDB database. All SQL queries (SQLServer2008) perform some type of data operation such as selecting data, inserting/updating data, or creating data objects such as SQL databases and SQL tables.

- **SQLQuery.RKN.contigs.sql**

```
SELECT * FROM [MA.contigs]
UNION ALL
SELECT * FROM [MC.contigs]
UNION ALL
SELECT * FROM [MH.contigs]
UNION ALL
SELECT * FROM [MI.contigs]
UNION ALL
SELECT * FROM [MJ.contigs]
UNION ALL
SELECT * FROM [MP.contigs]
GO
```

- **UNION all six species**

```
SELECT * FROM [MA.contigs]
UNION ALL
SELECT * FROM [MC.contigs]
UNION ALL
SELECT * FROM [MH.contigs]
UNION ALL
```

```
SELECT * FROM [MI.contigs]
UNION ALL
SELECT * FROM [MJ.contigs]
UNION ALL
SELECT * FROM [MP.contigs]
GO

SELECT * FROM MA_blastp_besthits
UNION ALL
SELECT * FROM MC_blastp_besthits
union ALL
SELECT * FROM MH_blastp_besthits
UNION ALL
SELECT * FROM MI_blastp_besthits
UNION ALL
SELECT * FROM MJ_blastp_besthits
UNION ALL
SELECT * FROM MP_blastp_besthits
ORDER BY [Contige]
GO
UPDATE MJ_blastp_besthits
SET E_value = REPLACE(E_value,',','')
GO
ALTER TABLE MJ_blastp_besthits
ALTER COLUMN E_value float (30) NOT NULL
GO
SELECT distinct b.[Contigs]
```

```
,b.[Score]
,b.[E-value]
,b.[GeneSeqName]
,b.[C#elegans]
,g.[WBGeneID]
,b.[Locus]
,b.[Protein]
,b.[Status]
,b.[UniProt]
,b.[Protein_ID]
,g.[WBRNAi]
,g.[WBPhenotype]
,g.[WBRNAi1]
,g.[WBPhenotype1]
,g.[WBRNAi2]
,g.[WBPhenotype2]
,g.[WBRNAi3]
,g.[WBPhenotype3]
,g.[WBRNAi4]
FROM [RKN_data].[dbo].[RKN_blastp_besthits]AS
b,[RKN_data].[dbo].[RKN_wormbase_batch_genes] AS g
WHERE b.[WBGeneID] = g.[WBGeneID]
GO
```

Bibliography

1. Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., ... & Rosso, M. N. (2008). Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nature biotechnology, 26(8), 909-915.
2. Alkharouf, N. W., Klink, V. P., & Matthews, B. F. (2007). Identification of Heterodera glycines (soybean cyst nematode [SCN]) cDNA sequences with high identity to those of Caenorhabditis elegans having lethal mutant or RNAi phenotypes. Experimental parasitology, 115(3), 247-258.
3. ATAWONG, A., HASEGAWA, M., & KODAMA, O. (2002). Biosynthesis of rice phytoalexin: enzymatic conversion of 3β-hydroxy-9β-pimara-7, 15-dien-19, 6β-olide to momilactone A. Bioscience, biotechnology, and biochemistry, 66(3), 566-570.
4. Bennett D (2010), Root-knot nematodes in soybeans. Delta Farm Press. Apr. 10, http://deltafarmpress.com/soybeans/root-knot-nematodes-soybeans.
5. Bybd Jr, D. W., Kirkpatrick, T., & Barker, K. R. (1983). An improved technique for clearing and staining plant tissues for detection of nematodes. Journal of Nematology, 15(1), 142.
6. Caillaud, M. C., Dubreuil, G., Quentin, M., Perfus-Barbeoch, L., Lecomte, P., de Almeida Engler, J., ... & Favery, B. (2008). Root-knot nematodes manipulate plant cell functions during a compatible interaction. Journal of plant physiology, 165(1), 104-113.

7. Collakova, E., & DellaPenna, D. (2003). The role of homogentisate phytyltransferase and other tocopherol pathway enzymes in the regulation of tocopherol synthesis during abiotic stress. Plant physiology, 133(2), 930-940.
8. Clauß, K., Baumert, A., Nimtz, M., Milkowski, C., & Strack, D. (2008). Role of a GDSL lipase like protein as sinapine esterase in Brassicaceae. The Plant Journal, 53(5), 802-813.
9. David, M. B., Valerie, M. W., Pierre, A., James, M., Etienne, G. J., Philippe, C., and Charles, H. O. (2009), the Genomes of Root-Knot Nematodes. Annual Review Phytopathol, (47), 333-351.
10. Davis, E. L., Hussey, R. S., & Baum, T. J. (2004). Getting to the roots of parasitism by nematodes. TRENDS in Parasitology, 20(3), 134-141.
11. Dawar, S. H. A. H. N. A. Z., Sattar, A., & Zaki, M. J. (2008). Seed dressing with biocontrol agents and nematicides for the control of root knot nematode on sunflower and okra. Pakistani Journal of Botany, 40, 2683-2691. http://www.pakbs.org/pjbot/PDFs/40(6)/PJB40(6)2683.pdf
12. Eisenback, J. D., & Triantaphyllou, H. H. (1991). Root-knot nematodes: Meloidogyne species and races. Manual of agricultural nematology, 191-274.
13. Endo, B. Y. (1964). Penetration and development of Heterodera glycines in soybean roots and related anatomical changes. Phytopath, 54, 79-88.
14. Endo, B. Y. (1965). Histological responses of resistant and susceptible soybean varieties, and backcross progeny to entry and development of Heterodera glycines. Phytopathology 55:375–381.

15. Endo, B. Y. (1991). Ultrastructure of initial responses of susceptible and resistant soybean roots to infection by Heterodera glycines. Revue de Ne´matologie 14:73–94
16. Eycken, W., Engler, J., Inzé, D., Montagu, M., & Gheysen, G. (1996). A molecular study of root-knot nematode-induced feeding sites. The Plant Journal, 9(1), 45-54.
17. Gibson, G. (2003). Microarray Analysis. PloS Biol 1(1): e15. Doi: 10.1371 / journal. pbio.0000015
18. Graham, R. M. (2005). Archives of Biochemistry and Biophysics, 433 (pp. 117–128).
19. Grundler, F. M., & Böckenhoff, A. (1997). Physiology of nematode feeding and feeding sites. In Cellular and Molecular Aspects of Plant-Nematode Interactions (pp. 107-119). Springer Netherlands.
20. Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K., Hannick, L. I., ... & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic acids research, 31(19), 5654-5666.
21. Hermsmeier, D., Mazarei, M., & Baum, T. J. (1998). Differential display analysis of the early compatible interaction between soybean and the soybean cyst nematode. Molecular plant-microbe interactions, 11(12), 1258-1263.
22. Hosseini, P., Tremblay, A., Matthews, B. F., and Alkharouf, N. W. (2012). MAPT and PAICE: Tools for time series and single time point transcriptionist visualization and knowledge discovery. Bioinformation, 8(6): 287-289.

23. Ibrahim, H. M., Alkharouf, N. W., Meyer, S. L., Aly, M. A., Gamal El-Din, A. E. K. Y., Hussein, E. H., & Matthews, B. F. (2011). Post-transcriptional gene silencing of root-knot nematode in transformed soybean roots. Experimental parasitology, 127(1), 90-99.
24. Ibrahim, H., Hosseini, P., Alkharouf, N., Hussein, E., El-Din, A. E. K. G., Aly, M., & Matthews, B. (2011). Analysis of Gene expression in soybean (Glycine max) roots in response to the root knot nematode Meloidogyne incognita using microarrays and KEGG pathways. BMC genomics, 12(1), 220.
25. Jung, C., & Wyss, U. (1999). New approaches to control plant parasitic nematodes. Applied Microbiology and Biotechnology, 51(4), 439-446.
26. Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., ... & Bohnert, H. J. (2001). Gene expression profiles during the initial phase of salt stress in rice. The Plant Cell Online, 13(4), 889-905.
27. Kim, Y. H., Riggs, R. D., & Kim, K. S. (1987). Structural changes associated with resistance of soybean to Heterodera glycines. Journal of nematology,19 (2), 177.
28. Klink, V. P., Overall, C. C., Alkharouf, N. W., MacDonald, M. H., & Matthews, B. F. (2007). A time-course comparative microarray analysis of an incompatible and compatible response by Glycine max (soybean) to Heterodera glycines (soybean cyst nematode) infection. Planta, 226(6), 1423-1447.
29. Klink, V. P., Overall, C. C., Alkharouf, N. W., MacDonald, M. H., & Matthews, B. F. (2007). Laser capture microdissection (LCM) and comparative microarray expression analysis of syncytial cells isolated from incompatible and compatible

soybean (Glycine max) roots infected by the soybean cyst nematode (Heterodera glycines). Planta, 226(6), 1389-1409.

30. Klink, V. P., Kim, K. H., Martins, V., MacDonald, M. H., Beard, H. S., Alkharouf, N. W., ... & Matthews, B. F. (2009). A correlation between host-mediated expressions of parasite genes as tandem inverted repeats and abrogation of development of female Heterodera glycines cyst formation during infection of Glycine max. Planta, 230(1), 53-71.
31. Lamb, C., & Dixon, R. A. (1997). The oxidative burst in plant disease resistance. Annual review of plant biology, 48(1), 251-275.
32. Lindahl, M., Spetea, C., Hundal, T., Oppenheim, A. B., Adam, Z., & Andersson, B. (2000). The thylakoid FtsH protease plays a role in the light-induced turnover of the photosystem II D1 protein. The Plant Cell Online, 12(3), 419-431.
33. Mahalingam, R., & Skorupska, H. T. (1996). Cytological expression of early response to infection by Heterodera glycines Ichinohe in resistant PI 437654 soybean. Genome, 39(5), 986-998.
34. Martin, J., Abubucker, S., Wylie, T., Yin, Y., Wang, Z., & Mitreva, M. (2009). Nematode.net updates 2008: improvements enabling more efficient data mining and comparative nematode genomics. Nucleic acids research, 37(suppl 1), D571-D578.
35. Meyer, S. L., Massoud, S. I., Chitwood, D. J., & Roberts, D. P. (2000). Evaluation of Trichoderma virens and Burkholderia cepacia for antagonistic activity against root-knot nematode, Meloidogyne incognita. Nematology, 2(8), 871-879.

36. Miyawaki, K., Tarkowski, P., Matsumoto-Kitano, M., Kato, T., Sato, S., Tarkowska, D., ... & Kakimoto, T. (2006). Roles of Arabidopsis ATP/ADP isopentenyltransferases and tRNA isopentenyltransferases in cytokinin biosynthesis. Proceedings of the National Academy of Sciences, 103(44), 16598-16603.

37. Newman, J. W., Morisseau, C., & Hammock, B. D. (2005). Epoxide hydrolases: their roles and interactions with lipid metabolism. Progress in lipid research, 44(1), 1-51.

38. Niebel, A., Heungens, K., Barthels, N., Inzé, D., Montagu, M. V., & Gheysen, G. (1995). Characterization of a pathogen-induced potato catalase and its systemic expression upon nematode and bacterial infection. MPMI-Molecular Plant Microbe Interactions, 8(3), 371-378.

39. Olsen, W. M. (2000). Root-knot Nematode. Collage of Agriculture and Life Science, Arizona University (2000), http://ag.arizona.edu/pubs/diseases/az1187

40. Platt, H. M. (1994). The phylogenetic systematics of free-living nematodes. The Ray Society, London, 383.

41. Sarkissian, C. N., & Gámez, A. (2005). Phenylalanine ammonia lyase, enzyme substitution therapy for phenylketonuria, where are we now?. Molecular genetics and metabolism, 86, 22-26.

42. Stirling, G. R., Stanton, J. M., & Marshall, J. W. (1992). The importance of plant-parasitic nematodes to Australian and New Zealand agriculture.Australasian plant pathology, 21(3), 104-115.

http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=12824352.

43. Vaghchhipawala, Z., Bassüner, R., Clayton, K., Lewers, K., Shoemaker, R., & Mackenzie, S. (2001). Modulations in gene expression and mapping of genes associated with cyst nematode infection of soybean. Molecular plant-microbe interactions, 14(1), 42-54.

44. Widmer, T. L., Ludwig, J. W., & Abawi, G. S. (1999). The northern root-knot nematode on carrot, lettuce, and onion in New York. New York State Agricultural Experiment Station.

http://vegetablemdonline.ppath.cornell.edu/factsheets/RootKnotNematode.htm

45. Wrather, J. A., & Koenning, S. R. (2006). Estimates of disease effects on soybean yields in the United States 2003 to 2005. Journal of nematology, 38(2), 173-180.

46. Xie, Z. M., Zou, H. F., Lei, G., Wei, W., Zhou, Q. Y., Niu, C. F., & Chen, S. Y. (2009). Soybean trihelix transcription factors GmGT-2A and GmGT-2B improve plant tolerance to abiotic stresses in transgenic Arabidopsis. PLoS One, 4(9), e6898.

47. Zhang, X. D. (2011). Optimal high-throughput screening: practical experimental design and data analysis for genome-scale RNAi research. Cambridge University Press.

Curriculum Vitae (CV)

NAME: Ahmed Ismail

PERMANENT ADDRESS: 506 Castle Dr., Apt. E, Baltimore, MD 21212

DEGREE AND DATE TO BE CONFERRED:

- **Towson University, Maryland Graduation: Aug 2013**
  - Degree: Doctor of Science in Information Technology (Bioinformatics)
- **Towson University, Maryland Graduation: May 2009**
  - Degree: Master of Science in Applied Information Technology
- **Azhar University, Cairo, Egypt Graduation: May2002**
  - Degree: Bachelor of Science in Biotechnology

Major: Computer and Information Science

Minor(s): Bioinformatics

Professional publications:

- Ahmed Ismail and Nadim W. Alkharouf Bioinformatics tools and analysis to identify lethal genes in Root Knot Nematode. In Vitro Biology Journal 2011, Raleigh, North Carolina, USA
- Ahmed Ismail, Benjamin F. Matthews, Nadim W. Alkharouf "RKN Lethal DB: A database for the identification of Root Knot Nematode (Meloidogyne spp) candidate lethal genes. Bioinformation 2012; 8(19): 950-952 PMID:23144556 [PubMed] PMCID:PMC3488838

Professional positions held:

- **Towson University (Maryland, USA)**
  - Lecturer in the Department of Computer and Information Science
    September 2013 to present (by the date of writing the dissertation)
  - Adjunct Faculty in the Department of Computer and Information Science
    September 2011 to May 2013
  - Teaching Assistant in the Department of Computer and Information Science
    September 2010 to September 2011
  - Research Assistant in the Department of Computer and Information Science
    September 2008 to January 2010
- **Azhar University (Cairo, Egypt)**
  - Teaching Assistant May 2003 to September 2007